PROCEEDINGS

OF THE

ROYAL SOCIETY OF LONDON

SERIES A. MATHEMATICAL AND PHYSICAL SCIENCES

VOL CLXVI

LONDON
Printed and published for the Royal Society
By the Cambridge University Press
Bentley House, N.W. 1

16 June 1938

PRINTED IN GREAT BRITAIN BY
WALTER LEWIS, M.A.
AT THE CAMBRIDGE UNIVERSITY PRESS

CONTENTS

SERIES A VOL CLXVI

No. A 924. 4 May 1938

PAGE

On the diamagnetic susceptibility of bismuth. By M. Blackman . . 1

The measurement of particle size by the X-ray method. By F. W. Jones 16

The resistance of superconducting cylinders in a transverse magnetic field. By A. D. Misener 43

Progressive lightning. V. A comparison of photographic and electrical studies of the discharge process. By B. F. J. Schonland, D. B. Hodges and H. Collens. (Plates 1–3) 56

X-ray studies of the structure of hair, wool, and related fibres. IV. The molecular structure and elastic properties of the biological cells. By H. J. Woods. (Plates 4, 5) 76

The development of the spark discharge. By T. E. Allibone and J. M. Meek. (Plates 6–11) 97

Quantum theory of Einstein-Bose particles and nuclear interaction. By N. Kemmer 127

On the nuclear forces and the magnetic moments of the neutron and the proton. By H. Fröhlich, W. Heitler and N. Kemmer 154

No. A 925. 19 May 1938

The lift and moment on a flat plate in a stream of finite width. By T. H. Havelock, F.R.S. 178

Some experimental observations for longitude made by theodolite fitted with a shutter eyepiece. By J. de Graaff Hunter, F.R.S. . . . 197

The cascade theory of electronic showers. By L. Landau and G. Rumer 213

Polish on metals. By W. Cochrane. (Plates 12, 13) 228

The spectrum of thallium chloride. By H. G. Howell and N. Coulson. (Plate 14) 238

β-Transitions in a coulomb field. By F. Hoyle 249

PAGE

The electrical conductivity of thin metallic films. III. Alkali films with the properties of the normal metal. By A. C. B. Lovell 270

The dispersion formula for nuclear reactions. By P. L. Kapur and R. Peierls 277

A note on Guggenheim's theory of strictly regular binary liquid mixtures. By G. S. Rushbrooke 296

The dielectric polarization of a *n*-long chain ketone at constant volume and variable temperature. By Alex Müller 316

No. A 926. 3 June 1938

Diamagnetism of cadmium. By S. Ramachandra Rao and S. Sriraman 325

The photosensitivity of diphenylamine-*p*-diazonium sulphate by the method of photometric curves. By C. F. Goodeve and L. J. Wood 342

An X-ray study of the iron-nickel-aluminium ternary equilibrium diagram. By A. J. Bradley and A. Taylor 353

Atomic rearrangement process in the copper-gold alloy Cu_3Au. II. By F. W. Jones and C. Sykes. (Plate 15) 376

The statistical mechanics of condensing systems. By Max Born and Klaus Fuchs 391

An accurate determination of the range-distribution curve of the radioactive α-particles from ^{8}Li. By C. L. Smith and W. Y. Chang . . 415

On the theory of scattering of light. By Hans Mueller 425

No. A 927. 16 June 1938

Self-consistent field with exchange for potassium and argon. By D. R. Hartree, F.R.S., and W. Hartree 450

The problem of *n* bodies in general relativity theory. By Sir Arthur Eddington, F.R.S., and G. L. Clark 465

Measurements with a half-pitot tube. By G. I. Taylor, F.R.S. . . . 476

The energy loss of penetrating cosmic-ray particles in copper. By J. G. Wilson. (Plate 16) 482

PAGE

On the theory of heavy electrons and nuclear forces. By H. J. Bhabha . 501

Showers produced by the penetrating cosmic radiation. By W. Heitler . 529

The formation of helium molecules. By F. L. Arnot and Marjorie B. M'Ewen 543

Application of "reciprocity" to nuclei. By M. Born 552

The adsorption of vapours at plane surfaces of mica. II. Heats of adsorption and the structure of multimolecular films. By D. H. Bangham and S. Mosallam 558

The swelling of charcoal. Part V. The saturation and immersion expansions and the heat of wetting. By D. H. Bangham and R. I. Razouk . 572

Index 587

On the diamagnetic susceptibility of bismuth

By M. Blackman, Trinity College, Cambridge

(*Communicated by R. H. Fowler, F.R.S.—Received* 4 *October* 1937—
Revised 27 *January* 1938)

The diamagnetic susceptibility of free electrons was first treated by Landau (1930) who discussed the behaviour of electrons in a magnetic field and showed the motion along the field to be purely classical, while in a plane at right angles to this direction quantum conditions hold which are analogous to those of a two-dimensional oscillator. The investigation showed that in weak fields the diamagnetism had a third of the value of the spin paramagnetism.

This theory would not appear at first sight to have any immediate application to metals because the collision of the electrons with the lattice would appear to inhibit the spiral motion, and hence the discrete eigenvalues of the Landau treatment do not exist. Peierls (1933 *a*, *b*) had, however, shown that the Landau formula holds for metals in spite of this difficulty.

Under the assumption that the electron is tightly bound it was further shown that the energy of the electron in a magnetic field can be put into the form

$$E = E_1 + \epsilon(k), \tag{1}$$

where E_1 and ϵ will in general be functions of the field strength and k is the wave vector whose components obey the commutation relations of the type

$$k_x k_y - k_y k_x = ieH_z/\hbar c. \tag{2}$$

In the case of bismuth there is a great deal of evidence showing that the number of electrons in one of the energy bands is very small; hence it was assumed by Peierls that only the lowest energy levels are important, in which case the function ϵ is needed only in the neighbourhood of its minimum. It was assumed that E could be written in the form

$$E = E_1 + \tfrac{1}{2}B(H)\,k^2. \tag{3}$$

In this form the energy differs from that for free electrons only in that the mass is replaced by $\hbar^2/B$ and that the zero of energy is changed. The solution of the eigenvalue problem follows now from the work of Landau

$$E = E_1 + (2n+1)\,\mu' H + Bp_z^2/2\hbar^2, \tag{4}$$

where $\mu' = eB/2\hbar c$. From this the free energy of the system can be calculated and the magnetic moment per unit volume follows from the relation

$$M = -\frac{1}{V}\frac{\partial F}{\partial H}. \tag{5}$$

The formulae were evaluated neglecting the dependence of E_1 and B on H. The results of the calculation of the magnetic moment as a function of the field strength for various temperatures were compared by Peierls with the experimental curves of de Haas and van Alphen (1930). Qualitatively the agreement was good, but it was not possible to obtain quantitative agreement.

This paper is an extension of the work of Peierls generalizing his treatment with the object of improving the quantitative side of the results and also of accounting for the finer features of the susceptibility curves found in the recent work of van Alphen (1933) and of Shoenberg and Uddin (1936).

I assume, first, a more general form for the function ϵ,

$$\epsilon = a_1 k_x^2 + a_2 k_y^2 + a_3 k_z^2, \tag{6}$$

i.e. the surfaces of constant energy in k-space are assumed to be ellipsoidal instead of spherical.

A function of the type (6) must show axial symmetry about the trigonal axis if it is to have the symmetry of the bismuth lattice. This, however, would lead to a magnetization curve which would be the same for each direction at right angles to the principal axis. Since the experimental curves show pronounced differences for directions along and at right angles to the binary axis, it must be concluded that the important part of the energy surface consists of several ellipsoids which separately have no trigonal symmetry but which transform into each other under a rotation of 120° about the trigonal axis. There must be at least three of these ellipsoids and in the following calculations we shall improve on those of Peierls mainly by assuming three ellipsoids instead of one.

The fact that the surfaces of constant energy contain several ellipsoids follows from the work of Jones (1934) on the Brillouin zones of bismuth. The following investigation shows that the ellipsoids needed to explain the de Haas-van Alphen effect are not identical with those found by Jones. This is no contradiction, as it will be shown that the electrons associated with our ellipsoids form less than 1 % of the total number of conduction electrons; hence this work does not conflict with that of Jones on the origin of the main part of the susceptibility along the different axes.

In the picture given by Jones the surfaces of discontinuity of the energy contain only the planes $(1\bar{1}0)$, $(\overline{22}1)$ and (221). Actually the planes of discontinuity such as (131) must also be present even if one assumes (with Jones) that the corresponding Fourier coefficients of the potential are negligibly small; for a reflexion by (221) and a subsequent reflexion by $(1\bar{1}0)$ is equivalent to one reflexion by (131), and such double reflexions do occur in second approximation. The planes (131), etc., cut across the main zone discussed by Jones forming new corners and edges. It is likely that the electronic states now under discussion are to be found near one of these new corners, but it seems difficult to localize these states without knowing more about the energy surfaces.

1. REDUCTION OF THE ENERGY SURFACE TO THE ISOTROPIC FORM

Write (6) in the form

$$E = \frac{1}{2m}(\alpha_1 p_x^2 + \alpha_2 p_y^2 + \alpha_3 p_z^2), \tag{7}$$

where α_1, α_2, α_3 are dimensionless constants, p_x, p_y, p_z are components of the momentum of the electron and m is its mass. Then the same relation (7) holds in a magnetic field except that p_x, p_y, p_z, x, y, z now satisfy the following commutation relations:

$$[p_x p_y] = ie\hbar H_z/c, \quad [p_x x] = i\hbar, \tag{8}$$

with similar relations for p_y, p_z, y, z. The energy can be expressed in the "spherical" form by putting $p'_x = l_x p_x$, $x' = \lambda_x x$ with similar relations for p_y, p_z, y, z. Hence (7) becomes

$$E = \frac{1}{2m}\left(\frac{\alpha_1}{l_x^2}p_x'^2 + \frac{\alpha_2}{l_y^2}p_y'^2 + \frac{\alpha_3}{l_z^2}p_z'^2\right) = \frac{1}{2m}p'^2 \tag{9}$$

if we put $\alpha_1 = l_x^2$, $\alpha_2 = l_y^2$, $\alpha_3 = l_z^2$. In the new variables the commutation relations become

$$[p'_x p'_y] = ie\hbar l_x l_y H_z/c = ie\hbar H_z(\alpha_1\alpha_2)^{\frac{1}{2}}/c = ie\hbar H'_z/c \tag{10}$$

and

$$[p'_x x'] = i\hbar_x \lambda_x = i\hbar,$$

with a similar form for the other relations, if

$$\lambda_x = 1/l_x, \quad \lambda_y = 1/l_y, \quad \lambda_z = 1/l_z,$$

$$H'_z = (\alpha_1\alpha_2)^{\frac{1}{2}} H_z, \quad H'_y = (\alpha_1\alpha_3)^{\frac{1}{2}} H_y, \quad H'_x = (\alpha_2\alpha_3)^{\frac{1}{2}} H_x.$$

Hence a volume element V' in the new co-ordinate system is related to that in the old co-ordinate system in the following way:

$$V' = dx'dy'dz' = \lambda_x \lambda_y \lambda_z V = V(\alpha_1 \alpha_2 \alpha_3)^{-\frac{1}{2}}. \tag{11}$$

The energy expression (7) can therefore be reduced to the "spherical" form by replacing the components of the field H by those of H' and by substituting the co-ordinates x', y', z' for x, y, z. The expression for H' is given by

$$H' = (\alpha_1 \alpha_2 H_z^2 + \alpha_3 \alpha_1 H_y^2 + \alpha_2 \alpha_3 H_x^2)^{\frac{1}{2}}. \tag{12}$$

In the following we require a number of special cases of the above formula. Two typical cases occur when the field is (*a*) parallel to one axis (z-axis), (*b*) in a plane containing the x- and y-axis and making an angle of 60° with the x-axis.

$$\left.\begin{aligned} &(a)\ H_z = H,\ H_x = 0,\ H_y = 0; \quad H' = H(\alpha_1 \alpha_2)^{\frac{1}{2}}. \\ &(b)\ H_z = 0, \quad H_x^2 = H^2 \cos^2 60^\circ,\ H_y^2 = H^2 \sin^2 60^\circ; \\ &\qquad\qquad H' = H[(3\alpha_1 \alpha_3 + \alpha_2 \alpha_3)/4]^{\frac{1}{2}}. \end{aligned}\right\} \tag{13}$$

2. CALCULATION OF THE DIAMAGNETIC MOMENT FOR THE ISOTROPIC CASE

The free energy of a system of electrons is given by the relation

$$F = -kT \sum_i \log(1 + e^{-(E_i - E_0)/kT}) + NE_0, \tag{14}$$

where N is the total number of electrons, i one of the possible states of motion, and E_i the energy for state i. E_0 is determined by the condition

$$\sum_i \frac{1}{1 + e^{(E_i - E_0)/kT}} = N. \tag{15}$$

Introducing the number of states of motion per unit volume $Z(E, H)$ which have a value of the energy below E, (14) and (15) become

$$\left.\begin{aligned} F &= -kTV \int_{-\infty}^{+\infty} \frac{dZ}{dE} \log(1 + e^{-(E_i - E_0)/kT})\, dE + NE_0, \\ N &= V \int_{-\infty}^{+\infty} \frac{dZ}{dE} \frac{1}{1 + e^{(E - E_0)/kT}}\, dE. \end{aligned}\right\}. \tag{16}$$

Integrating by parts this can be put into the form

$$\left.\begin{aligned} F &= V\int_{-\infty}^{+\infty} Z(E,H)\, e^{-(E-E_0)/kT}(1+e^{-(E-E_0)/kT})^{-1}\,dE + NE_0, \\ N &= \frac{V}{kT}\int_{-\infty}^{+\infty} Z(E,H)\,(1+e^{(E-E_0)/kT})^{-1}\,(1+e^{-(E-E_0)/kT})^{-1}\,dE. \end{aligned}\right\} \quad (17)$$

To determine Z we separate the motion into that parallel to the field and that perpendicular to it, the energy being E_1 and E_2 in the two cases.

Parallel to the field we have free motion, hence $E_1 = p^2/2m$, p taking the values $p = hn/L$ ($n = 0, \pm 1, \pm 2, \pm 3, \ldots$), where L is the length of the piece of metal. Hence there are $2L(2mE_1)^{\frac{1}{2}}/h$ states with "translational" energy below E_1. For the motion perpendicular to the field $E_2 = (2l+1)\,\mu H$, where $\mu = e\hbar/2mc$ and $l = 0, 1, 2, \ldots$. Each of the above energy states is degenerate and belongs to $eHD/2\pi\hbar c$ states of the two-dimensional motion, where D is the cross-section of the piece of metal. Besides this we must take account of the electronic spin which means that we must double the density of states as calculated above. The formula for Z is given by

$$VZ(E,H) = \frac{eH(2m)^{\frac{1}{2}}V}{\pi^2\hbar^2 c}\sum_l [E-(2l+1)\,\mu H]^{\frac{1}{2}}, \quad (18)$$

where $V = DL$. Hence

$$Z\,(E,H) = AH\,\Sigma\,[E-(2l+1)\,\mu H]^{\frac{1}{2}},$$

with $A = l(2m)^{\frac{1}{2}}/\pi^2\hbar^2 c$. Putting this value of Z into (17) and integrating by parts, we obtain the following expression for the moment per unit volume:

$$M = -\frac{1}{V}\frac{\partial F}{\partial H} = \frac{A}{3kT}\int dE \sum_l \frac{[E-(2l+1)\,\mu H]^{\frac{1}{2}}\,[2E-5(2l+1)\,\mu H]}{(1+e^{(E-E_0)/kT})\,(1+e^{-(E-E_0)/kT})}. \quad (19)$$

Introducing the new variables $\epsilon = E/\mu H$, $\epsilon_0 = E_0/\mu H$, $kT/\mu H = \theta$ we have

$$M = \frac{\mu^{\frac{5}{2}}H^{\frac{5}{2}}A}{3E_0}\int \varphi(\epsilon)\,\phi\!\left(\frac{\epsilon-\epsilon_0}{\theta}\right)d\epsilon, \quad (20)$$

where $$\varphi(\epsilon) = \sum_l [\epsilon-(2l+1)]^{\frac{1}{2}}\,[2\epsilon-5(2l+1)],$$

$$\phi\!\left(\frac{\epsilon-\epsilon_0}{\theta}\right) = \frac{E_0}{kT}\,\frac{1}{(1+e^{(\epsilon-\epsilon_0)/\theta})\,(1+e^{-(\epsilon-\epsilon_0)/\theta})}.$$

ϕ is a function which has the property that $\int_{-\infty}^{+\infty}\phi\!\left(\frac{\epsilon-\epsilon_0}{\theta}\right)d\epsilon = \epsilon_0$ and which has appreciable values only if the argument is numerically of order unity or less. Hence the integral in (20) has the effect of averaging the function φ over an interval of the order θ about ϵ_0.

At $T = 0$ (20) reduces to the expression

$$M = (\mu H)^{\frac{3}{2}} \varphi(\epsilon_0) A/3. \tag{21}$$

In the same way we can treat the expression for N and obtain

$$\frac{N}{V} = \frac{A\mu^{\frac{5}{2}}H^{\frac{5}{2}}}{E_0} \int d\epsilon\, \psi(\epsilon)\, \phi\left(\frac{\epsilon-\epsilon_0}{\theta}\right), \tag{22}$$

where $\psi(\epsilon) = \Sigma\,[\epsilon-(2l+1)]^{\frac{1}{2}}$. At $T = 0$ this becomes

$$\frac{N}{V} = A\mu^{\frac{3}{2}}H^{\frac{3}{2}}\psi(\epsilon_0). \tag{23}$$

Formula (20) is that given in Peierls' paper*

$$M = \frac{e(2m)^{\frac{1}{2}}\, E_0^{\frac{3}{2}}}{\pi^2\hbar^2 c}\frac{1}{3\epsilon_0^{\frac{3}{2}}}\int \varphi(\epsilon)\, \phi\left(\frac{\epsilon-\epsilon_0}{\theta}\right) d\epsilon = \frac{e(2m)^{\frac{1}{2}}\, E_0^{\frac{3}{2}}}{\pi^2\hbar^2 c}\, \sigma\left(\frac{\mu H}{kT}, \frac{kT}{E_0}\right),$$

where σ is the function given in diagram form in the paper.

3. THE DIAMAGNETIC MOMENT FOR THE ANISOTROPIC CASE

In the further applications we shall concern ourselves only with the case of $T = 0$. The effect of temperature will be briefly considered at the appropriate stages.

I shall first consider the modifications to be introduced into § 1 to take account of the fact that the energy has the "ellipsoidal" instead of the "spherical" form assumed there. We have seen that this can be accounted for by replacing V by V' and H by H', where V' and H' are given by (11) and (12). We can write the free energy in the form (see (17))

$$F = -V'G(E, H', T). \tag{24}$$

Hence
$$M = -\frac{1}{V}\frac{\partial F}{\partial H} = -\frac{V'}{V}\cdot\frac{H'}{H}\cdot\frac{\partial G}{\partial H'}.$$

The rest of the calculation can now be performed exactly as in the previous section, introducing the new variable $\epsilon_0' = E_0/\mu H'$ at the appropriate stage. We obtain finally

$$\begin{aligned} M &= \frac{H'}{H}\frac{V'}{V}(\mu H')^{\frac{5}{2}}\frac{A}{3E_0}\varphi(\epsilon_0')\,\epsilon_0' \\ &= \frac{V'}{V}\left(\frac{H'}{H}\right)^{\frac{5}{2}}(\mu H)^{\frac{3}{2}}\frac{A}{3}\varphi(\epsilon_0'). \end{aligned} \tag{25}$$

---

* The contents of this section were taken from unpublished work by Peierls.

Now V'/V and H'/H are functions of α_1, α_2, α_3 alone. Hence the only difference as compared with (20) is that there is an extra coefficient and also H' instead of H in the expression for φ. The calculation of the formula for N proceeds on similar lines and yields

$$\frac{N}{V} = \frac{A}{\mu}(\mu H)^{\frac{3}{2}}\frac{V'}{V}\left(\frac{H'}{H}\right)^{\frac{3}{2}}\psi(\epsilon_0'). \tag{26}$$

We see that the coefficient is different from that for M by a factor H'/H.

For the purposes of the applications I propose to consider the three cases mentioned in § 1 where the field is in specified directions. We take three ellipsoids which are so arranged that they go over into one another by rotation through 120°. Then taking the field at right angles to the principal axis and along a binary axis, we see that for one ellipsoid the field is directed along one axis while for the other two it makes angles of 60° and 120° with the corresponding axes. Hence for one ellipsoid $H' = H(\alpha_1\alpha_3)^{\frac{1}{2}}$, while for the other two $H' = H[(\alpha_2\alpha_3 + 3\alpha_1\alpha_3)/4]^{\frac{1}{2}}$. In the second case (at right angles to the binary axis) all we need to do is to interchange α_1 and α_2. When finding the value of the total moment we must add the effects of the ellipsoids; we can also form the expression for the total number of electrons belonging to the three ellipsoids by adding the corresponding three terms. We have

$$\left.\begin{aligned}M &= \frac{\mu^{\frac{3}{2}}e(2m)^{\frac{1}{2}}H^{\frac{3}{2}}}{3\pi^2\hbar^2c(\alpha_1\alpha_2\alpha_3)^{\frac{1}{2}}}\\ &\quad\times\left[(\alpha_1\alpha_3)^{\frac{1}{4}}\varphi\left(\frac{E_0}{\mu H(\alpha_1\alpha_3)^{\frac{1}{2}}}\right)+2(\tfrac{3}{4}\alpha_2\alpha_3+\tfrac{1}{4}\alpha_1\alpha_3)^{\frac{1}{4}}\varphi\left(\frac{E_0}{\mu H(\frac{3}{4}\alpha_2\alpha_3+\frac{1}{4}\alpha_1\alpha_3)^{\frac{1}{2}}}\right)\right],\\ \frac{N}{V} &= \frac{\mu^{\frac{1}{2}}e(2m)^{\frac{1}{2}}H^{\frac{3}{2}}}{3\pi^2\hbar^2c(\alpha_1\alpha_2\alpha_3)^{\frac{1}{2}}}\\ &\quad\times\left[(\alpha_1\alpha_3)^{\frac{3}{4}}\psi\left(\frac{E_0}{\mu H(\alpha_1\alpha_3)^{\frac{1}{2}}}\right)+2(\tfrac{3}{4}\alpha_2\alpha_3+\tfrac{1}{4}\alpha_1\alpha_3)^{\frac{3}{4}}\psi\left(\frac{E_0}{\mu H(\frac{3}{4}\alpha_2\alpha_3+\frac{1}{4}\alpha_1\alpha_3)^{\frac{1}{2}}}\right)\right]\end{aligned}\right\} \tag{27}$$

for the one case while the other is obtained by interchanging α_1 and α_2. Writing (27) in the form

$$\left.\begin{aligned}M &= \frac{\mu^{\frac{3}{2}}e(2m)^{\frac{1}{2}}H^{\frac{3}{2}}(\alpha_1\alpha_3)^{\frac{1}{4}}}{3\pi^2\hbar^2c(\alpha_1\alpha_2\alpha_3)^{\frac{1}{2}}}\Phi,\\ \frac{N}{V} &= \frac{\mu^{\frac{1}{2}}e(2m)^{\frac{1}{2}}H^{\frac{3}{2}}(\alpha_1\alpha_3)^{\frac{3}{4}}}{3\pi^2\hbar^2c(\alpha_1\alpha_2\alpha_3)^{\frac{1}{2}}}\Psi,\end{aligned}\right\} \tag{28 a}$$

with
$$\Phi = \varphi\left(\frac{E_0}{\mu H(\alpha_1\alpha_3)^{\frac{1}{2}}}\right) + 2\left(\frac{3\alpha_2}{4\alpha_1}+\frac{1}{4}\right)^{\frac{3}{4}} \varphi\left(\frac{E_0}{\mu H(\frac{3}{4}\alpha_2\alpha_3+\frac{1}{4}\alpha_1\alpha_3)^{\frac{1}{2}}}\right),$$
$$\Psi = \psi\left(\frac{E_0}{\mu H(\alpha_1\alpha_3)^{\frac{1}{2}}}\right) + 2\left(\frac{3\alpha_2}{4\alpha_1}+\frac{1}{4}\right)^{\frac{3}{4}} \psi\left(\frac{E_0}{\mu H(\frac{3}{4}\alpha_2\alpha_3+\frac{1}{4}\alpha_1\alpha_3)^{\frac{1}{2}}}\right),$$

it follows that
$$\left.\begin{aligned} M &= \kappa\Phi/\Psi, \\ H &= \kappa'\Psi^{-\frac{2}{3}}, \end{aligned}\right\} \tag{28b}$$

where κ and κ' can be determined from (28 a). If we make the assumption that N is a fixed constant, κ and κ' are constants and we can hence plot Φ/Ψ against $\Psi^{-\frac{2}{3}}$ if we wish to find how M varies with H.

In actual practice it is advantageous to plot $\chi = M/H$ against H, as this is more sensitive to small fluctuations and it is with these fluctuations that we are concerned.

To complete the argument at this stage we have still to discuss the formulae for the case where H is in the direction of the principal axis. This is particularly simple, as for all three ellipsoids $H' = H(\alpha_1\alpha_2)^{\frac{1}{2}}$, hence

$$\left.\begin{aligned} M &= \frac{3\mu^{\frac{1}{2}}e(2m)^{\frac{1}{2}}H^{\frac{3}{2}}}{\pi^2\hbar^2c(\alpha_1\alpha_2\alpha_3)^{\frac{1}{2}}}(\alpha_1\alpha_2)^{\frac{3}{4}}\varphi\left(\frac{E_0}{\mu H(\alpha_1\alpha_2)^{\frac{1}{2}}}\right), \\ \frac{N}{V} &= \frac{3\mu^{\frac{1}{2}}e(2m)^{\frac{1}{2}}H^{\frac{3}{2}}}{\pi^2\hbar^2c(\alpha_1\alpha_2\alpha_3)^{\frac{1}{2}}}(\alpha_1\alpha_2)^{\frac{3}{4}}\psi\left(\frac{E_0}{\mu H(\alpha_1\alpha_2)^{\frac{1}{2}}}\right). \end{aligned}\right\} \tag{29}$$

Here the form of the curves of M against H are identical with those for the "spherical" case except for scale changes.

This case is effectively the one which has been studied by Peierls (1933 a). The curve of M against H (at $T = 0$) consists of a large number of violent fluctuations which crowd together as H tends to zero, the final minimum as H increases occurring for $E_0 = 3\mu H$. After this the moment rises slowly to a constant value. This means that the susceptibility will tend to zero for large fields and hence the curve for the susceptibility will have an extra maximum otherwise the fluctuations will be the same as before.

As the temperature is raised the fluctuations are gradually smoothed out, those corresponding to low values of the field being the first to be affected. At high temperatures the susceptibility is independent of the field strength.

4. DISCUSSION OF THE EXPERIMENTAL CURVES

The experimental facts are contained in the dotted curves of fig. 1 and in figs. 2, 3, which are taken from the work of Shoenberg and Uddin (1936)

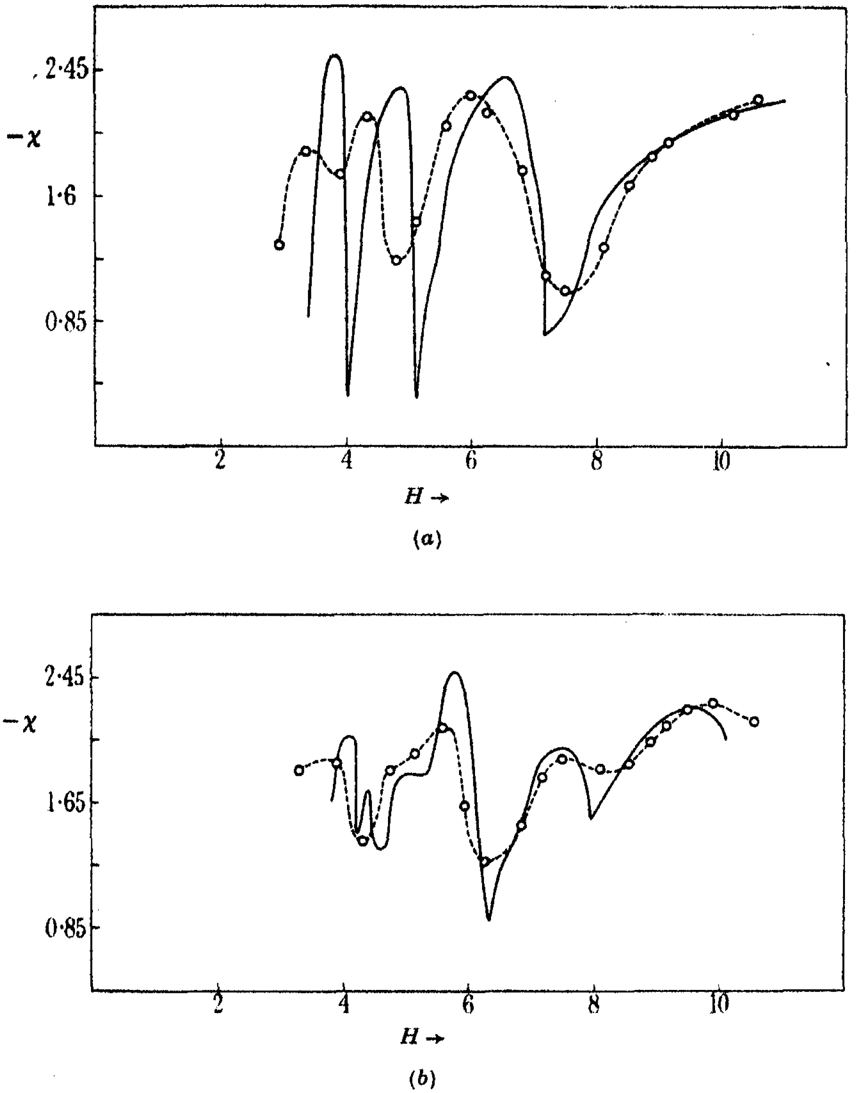


FIG. 1. Comparison of the experimental χ-H curves with the theoretical curves for $T = 0$ $(\alpha_1/\alpha_2 = 9{\cdot}8)$, (*a*) when the field is along the binary axis, (*b*) when the field is perpendicular to the binary axis. The experimental points are indicated in the curves. The susceptibility per unit volume is obtained by multiplying the ordinate by $9{\cdot}8 \times 10^{-6}$, the susceptibility per unit mass by multiplying by 10^{-6}. The unit for the abscissa is 10^3 gauss.

and de Haas and van Alphen (1933) respectively. The curves are those of the susceptibility[*] against the field strength, when the field is directed (*a*) along a binary axis of bismuth, (*b*) perpendicular to the binary axis.

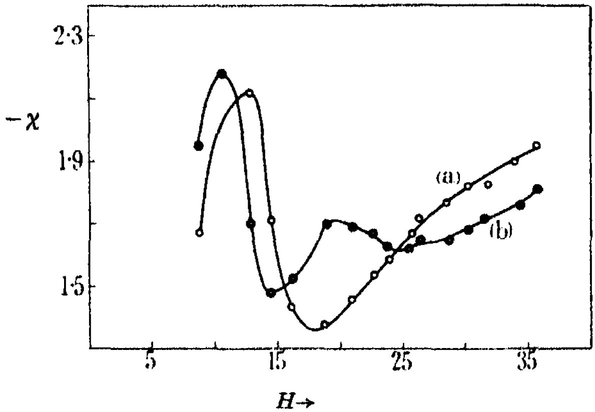


FIG. 2. Experimental χ-H values at $T = 14{\cdot}2^\circ$ K. for fields directed along (single circles) and perpendicular to the binary axis (double circles). The scale is the same as in fig. 1.

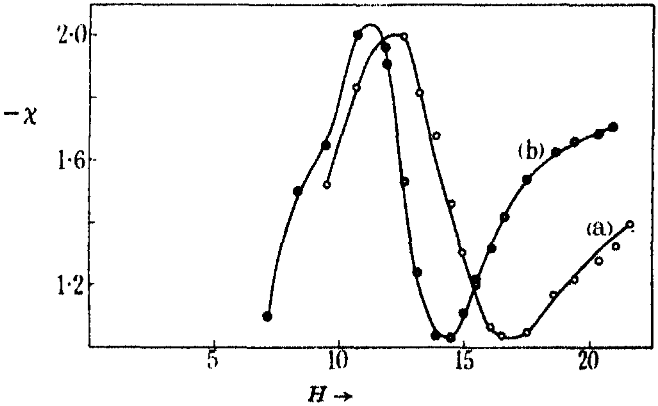


FIG. 3. Experimental χ-H curves for $T = 1{\cdot}86^\circ$ K. The scale is the same as in fig. 1.

Curve (*b*) shows secondary maxima which are not contained in curve (*a*), and is also shifted slightly with respect to it. There are also marked differences in the region of high fields, due to an additional maximum in curve (*a*) at about 20 kG.

---

[*] The susceptibility per unit volume is taken throughout.

The theory has to explain all these features as well as their absolute magnitude, and has also to show why there is no de Haas-van Alphen effect when the susceptibility is measured along the principal axis.

The measurements of Shoenberg and Uddin were carried from 4° K. upwards. The measurements of van Alphen show little change of the main maximum between 4 and 1·86° K.

The small difference in the values of the main maximum between 12 and 4° K. and the very slight change between 4 and 1·86° K. show that the curves have reached their final form by about 4° K. as far as the part of the curves lying above the main maximum (at about 10 kG.) is concerned. Since the moment is always negative, it must be concluded that a large part of the diamagnetism is due to electrons which do not show the de Haas-van Alphen effect.

5. COMPARISON OF THEORY AND EXPERIMENT

I shall first discuss briefly what happens when it is assumed that the number of electrons in the band responsible for the de Haas-van Alphen effect is fixed. The curves of χ as a function of H are given by (29a), (28b) and (20).* We see that it is only the relative values of the field strength and of the coefficients of Φ and Ψ which matter in fitting the curves. Hence we can express (27) in terms of the parameter α_1/α_2 alone, the other factors being important only when the absolute values are wanted. It is found that the two χ-H curves show a marked resemblance to the experimental ones only when α_1/α_2 has a value between 9 and 20. A typical curve for which $\alpha_1/\alpha_2 = 10$ is given in fig. 4. In order to see how

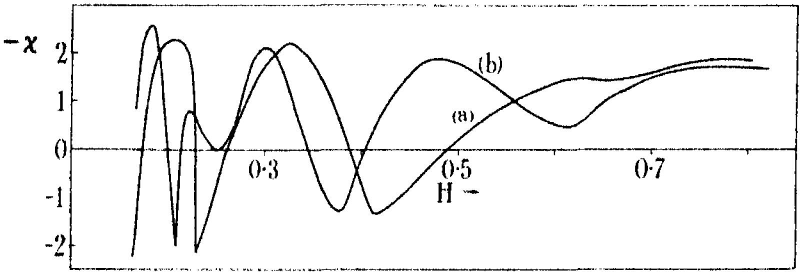


FIG. 4. Theoretical χ-H curves for $T = 0$ assuming a fixed number of electrons in the relevant band. The parameter α_1/α_2 has the value 10. The scale is arbitrary.

---

* The sums in (20) contain only a few terms and can hence be worked out to any desired accuracy.

the curves fit numerically one must fix the absolute values of the various parameters, and the zero value of the susceptibility. With a zero value $= 14\cdot2 \times 10^{-6}$ for the susceptiblity per unit volume, $E_0 = 3\cdot1 \times 10^{-14}$, $\alpha_1 = 13\cdot5$, $\alpha_2 = 1\cdot38$, $\alpha_3 = 8\cdot30$, the theoretical curves fit the experimental ones fairly well (the method of calculating these constants will be given below). The number of electrons in the band can be calculated from (27) and is found to be $1\cdot2 \times 10^{-5}$ which is about 1 % of the estimated number of conduction electrons per atom as estimated by Jones (1934).

The main part of the diamagnetism has been discussed by Jones (1934) who has shown this to be due to certain ellipsoids which form the most important part of the energy surface $E = E_0$ (the Fermi energy). The ellipsoids discussed above have a different shape and form a minor part of the energy surface not discussed by Jones. They are in general of less importance, as they contain a small fraction of the total number of conduction electrons and will be of importance only in special cases such as in the effect discussed here.

Because of the large number of other electrons, the assumption which we have made—that the number of electrons in the de Haas-van Alphen band is independent of the field strength—is not quite correct. We should assume that the total number of electrons in all the conducting bands together, should be a constant while the Fermi energy E_0 should be the same for all bands. Since, however, the important band contains so few electrons the value of E_0 will be determined practically entirely by the other bands. These do not show the de Haas-van Alphen effect, hence E_0 is independent of H.

I therefore modify the discussion given above by replacing the condition $N =$ const. by the condition $E_0 =$ const. Even if the number of electrons in the de Haas-van Alphen band is not negligible, we will obtain results intermediate between those discussed above and those given below.

With the assumption $E_0 =$ const. we can immediately write down the formulae for the magnetic moment associated with the various axes. The formulae are identical with (27) except that we may now replace H by $E_0/\mu\epsilon_0$, since $\epsilon_0 = E_0/\mu H$ and E_0 is now fixed. We have therefore

$$M = \frac{e(2m)^{\frac{1}{2}} E_0^{\frac{3}{2}} \epsilon_0^{-\frac{1}{2}}}{3\pi^2\hbar^2 c(\alpha_1\alpha_2\alpha_3)^{\frac{1}{2}}}$$

$$\times \left\{ (\alpha_1\alpha_3)^{\frac{1}{2}} \varphi\left(\frac{E_0}{\mu H(\alpha_1\alpha_3)^{\frac{1}{2}}}\right) + 2(\tfrac{3}{4}\alpha_2\alpha_3 + \tfrac{1}{4}\alpha_1\alpha_3)^{\frac{1}{2}} \varphi\left(\frac{E_0}{\mu H(\frac{3}{4}\alpha_2\alpha_3 + \frac{1}{4}\alpha_1\alpha_3)^{\frac{1}{2}}}\right)\right\} \quad (30)$$

for the moment with the field parallel to the binary axis, and have to interchange α_1 and α_2 in order to obtain the other case. This may be written as

$$M = C\phi(H_0/H), \tag{31}$$

where $H_0 = E_0/\mu(\alpha_1\alpha_2)^{\frac{1}{2}}$, $\beta = \left(\frac{3}{4}\frac{\alpha_2}{\alpha_1}+\frac{1}{4}\right)^{\frac{1}{2}}$, and $\phi(x) = \varphi(x)+2\beta^5\varphi(x/\beta^2)$. The moment associated with the principal axis is given by

$$M = \frac{e(2m)^{\frac{1}{2}}\,E_0^{\frac{3}{2}}\,\epsilon_0^{-\frac{3}{2}}}{\pi^2\hbar^2c(\alpha_1\alpha_2\alpha_3)^{\frac{1}{2}}}\,(\alpha_1\alpha_2)^{\frac{1}{2}}\,\varphi\left(\frac{E_0}{\mu H(\alpha_1\alpha_2)^{\frac{1}{2}}}\right). \tag{32}$$

One can again draw the χ-H curves for various values of the parameter α_1/α_2. It will be noted that these are very similar in many respects to the curves for the case N = const. It again appears that only when α_1/α_2 lies in the range 9–20 do the theoretical resemble the experimental values. We now proceed to the determination of the absolute values of the parameters involved in (31). E_0 is fixed by considering the variation of the experimental χ-H curves with temperature, and comparing them with the curves calculated* by Peierls (1933*b*) for three values of kT/E_0. The value determined in this way is $E_0 = 3{\cdot}1\times10^{-14}$. We now fix on a point on the experimental curve for which we can assume with sufficient accuracy that the limiting value of χ has been reached. This we take to be the main maximum on one curve at $H = 9{\cdot}8$ kG. We can now compare the moment at this point (less the zero value) with the theoretical value from (31). From the theoretical curve for $\alpha_1/\alpha_2 = 9{\cdot}8$ we can read off the value of ϵ_0 at this main maximum. This gives $E_0/\mu H(\alpha_1\alpha_3)^{\frac{1}{2}} = \epsilon_0 = 3{\cdot}3$. Hence $(\alpha_1\alpha_3)^{\frac{1}{2}} = 106$. We also know from (31) that

$$M = \frac{e(2m)^{\frac{1}{2}}\,E_0^{\frac{3}{2}}\,\epsilon_0^{-\frac{3}{2}}}{3\pi^2\hbar^2c(\alpha_1\alpha_2\alpha_3)^{\frac{1}{2}}}\,(\alpha_1\alpha_3)^{\frac{1}{2}}\,\Phi = 0{\cdot}0098(22{\cdot}6-\kappa_0),$$

where κ_0 is the zero value of the susceptibility in units of 10^{-6}. From this it follows that

$$\frac{1}{(\alpha_2)^{\frac{1}{2}}} = \frac{0{\cdot}0098(22{\cdot}6-\kappa_0)}{7{\cdot}4\times10^{-5}(\alpha_1\alpha_3)^{\frac{1}{2}}}.$$

In order to fix the value of κ_0 we need another equation. For this purpose we use the value of the susceptibility at high temperatures† which is

---

* The intensity (of the main maximum) was measured relative to the high temperature value in both cases in order to eliminate the "zero" term in the experimental curves.

† I.e. at such fields and temperature for which the susceptibility is independent of H.

$\kappa \times 10^{-6} = 0{\cdot}163 E_0^{\frac{1}{2}}\left(\frac{\alpha_1\alpha_3}{\alpha_2}\right)^{\frac{1}{2}}$.* This is also equal to the experimental value $\kappa' = 17{\cdot}5$ (at 77° K.) less the zero value κ_0 which will not change appreciably between 4 and 77° K. Since we know all the quantities except κ_0 we obtain its value directly: $\kappa_0 = 14{\cdot}5$. It follows that $\alpha_1 = 9{\cdot}8$, $\alpha_2 = 1{\cdot}0$, $\alpha_3 = 1{\cdot}1 \times 10^3$. The very high value of α_3 has as a consequence the effect of making the moment associated with the principal axis very small (less than 0·1 % of the total) as can be seen by comparing (32) with (30). This result fits in with the experimental fact that the de Haas-van Alphen effect is not found when the susceptibility is measured along the principal axis.

The way the theoretical curve fits the experimental ones is shown in fig. 1 (up to 10 kG.). It will be seen that the maxima and minima in each of the two curves are accounted for. The main difference lies (i) in the fact that the theoretical minima are sharp whereas those of the experimental curves are rounded off in a similar manner to the maxima, (ii) in that the maxima and minima of the theoretical curves are not sufficiently widely spaced.

With regard to the first point one must note that the effect of temperature will probably affect the sharp minima before the relatively flat maxima and also that in the experiments a small inhomogeneity of the field across the specimen would tend to flatten out the minima. The second point seems of more importance because a slight change of scale seems to be necessary for a better fit and this discrepancy also occurs when the curves for $N =$ const. are considered.

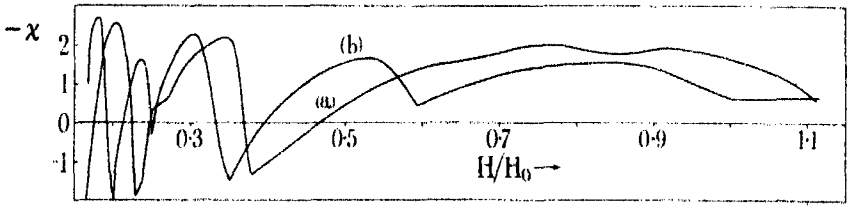


FIG. 5. Theoretical curves for $T = 0$ $(\alpha_1/\alpha_2 = 9{\cdot}8)$ extending to high values of the field. The scale factors are given in (31).

The experiments of Shoenberg and Uddin do not go beyond 11 kG. One can compare the theoretical curves for higher fields (fig. 5) with the experimental curves due to de Haas and van Alphen (figs. 2, 3) but as these are not complete at the lowest temperatures the theoretical curves were not specially fitted to the experimental ones. One can however see that there is an extra maximum in one of the theoretical curves at $H = 0{\cdot}54$ (this corresponds roughly to 20 kG.); this is also one of the most striking features of the experimental results at high fields. It can be concluded from the above comparison that the assumption of new ellipsoids in the energy surface accounts quite well for the features of the experimental curves. As pointed out in the introduction the existence of these ellipsoids, though not following from the work of Jones, is not in contradiction to it. The interesting feature is that the ellipsoids are surprisingly anisotropic and contain only a small fraction of the total number of conduction electrons. If the interpretation of the experimental curves is correct, they should approach the theoretical curves (figs. 1, 4) more closely as the temperature is lowered; in particular the minima should become much sharper. At higher fields further small anomalies are to be expected.

---

* This contains the effect of the three ellipsoids.

I should like to express my thanks to Dr R. Peierls, who suggested the problem, for his advice and interest. I am indebted to the Department of Scientific and Industrial Research for a Senior Research Award.

SUMMARY

An extension is made of the existing theory of the marked variation of the susceptibility of bismuth with field strength at low temperatures (known as the de Haas-van Alphen effect). The object of the investigation is to explain the finer features of the experiments, to account for the order of magnitude of the effect and for the dependence on the orientation of the binary axes. The extension consists in assuming a more general form for the relevant part of the surface of constant energy of the electrons. It is found possible to account for all the features of the susceptibility curves and also to fit the experimental curves fairly well. For this purpose one has to assume that the motion of the electrons is violently anisotropic and that only a small fraction of the conduction electrons contribute to the de Haas-van Alphen effect.

REFERENCES

van Alphen, P. M. 1933 "Enkele magnetische eigenschappen der metallen bij lage temperatuur." Amsterdam.

de Haas, W. J. and van Alphen, P. M. 1930 *Commun. phys. Lab. Univ. Leiden*, No. 212.

Jones, H. 1934 *Proc. Roy. Soc.* A, **147**, 396.

Landau, L. 1930 *Z. Phys.* **64**, 629.

Peierls, R. 1933 *a* *Z. Phys.* **80**, 763.

— 1933 *b* *Z. Phys.* **81**, 186.

Shoenberg, D. and Uddin, M. Z. 1936 *Proc. Roy. Soc.* A, **156**, 687.

The measurement of particle size by the X-ray method

By F. W. Jones, B.Sc.

Metropolitan-Vickers' Research Scholar at the University of Manchester

(*Communicated by W. L. Bragg, F.R.S.—Received* 23 *October* 1937)

I. Introduction

The broadening of the X-ray diffraction lines which occurs when the crystals composing the specimen are smaller than about 10^{-5} cm. edge length is well known. Since the discovery of this phenomenon by Scherrer in 1920 a fair amount of work has been done on the determination of the particle size from the breadth of the lines.

Scherrer (1920), Bragg (1933) and Seljakow (1925) have calculated the diffraction broadening for crystals of cubic form belonging to the cubic system. For parallel monochromatic radiation and a point specimen, they agree upon a formula

$$\beta_\chi = \frac{C\lambda}{t \cos \frac{1}{2}\chi}, \tag{1}$$

where β_χ = angular breadth of the line defined below,

t = edge length of cubic crystal,

λ = wave-length of X-radiation,

$\frac{1}{2}\chi = \theta$, the Bragg angle.

The values given for the constant C are 0·94, 0·89 and 0·92 respectively.

The most rigid theoretical treatment of the problem is that by Laue (1926), who has discussed the case of crystals of general parallelepiped form belonging to any crystal system, i.e. a crystal of edge lengths $m_1\mathbf{a}_1$, $m_2\mathbf{a}_2$, $m_3\mathbf{a}_3$, where the m_i's are numbers and the vectors $\mathbf{a}_i$ define the unit cell. For crystals of cubic form belonging to the cubic system, Laue's formula reduces to equation (1), the constant C having a value of 0·9. There are, however, two arithmetical errors in Laue's paper; if these are corrected C becomes 1·42. C involves another constant ω; by using a different method for evaluating ω we have obtained a value of 1·0 for C. Details of the calculation are given in Section V.

The constants given above are so nearly equal to unity, the value obtained in Section V, that for convenience we have defined a quantity ϵ, the "apparent particle size", by the formula

$$\beta_\chi = \frac{1\cdot 0\lambda}{\epsilon \cos \frac{1}{2}\chi}. \tag{2}$$

In the general case ϵ is a function of vectors in the reciprocal lattice and the edge lengths of the parallelepiped crystal. Its dependence on the particle shape will be considered later.

It must be realized that any such formula can give only an approximate estimate of particle size, and in fact little significance can be attached to a precise value of the constant C since it is not to be expected that the particles will be all of the same size. An accuracy of the order of 20 % can alone be aimed at.

The formulae of Scherrer, Bragg and Seljakow refer to a breadth of line defined as the angular width between the points where the intensity falls to half its maximum value, the "half value" breadth as it is called. Laue has used the following alternative definition. The line has maximum intensity at its centre and falls away on either side. The "breadth" is that of an ideal line which has uniform intensity equal to the maximum and a total intensity equal to that of the actual line, i.e.

$$\beta_\chi = \frac{\int I_{(\chi)}\, d\chi}{I_{(\max)}}. \tag{3}$$

This definition is more convenient for use in actual cases where the breadth of the line is influenced by the geometry of the apparatus.

The assumptions of parallel monochromatic radiation and a point specimen involved in the above formulae cannot be realized in practice.

Laue extended his calculations to a transparent specimen illuminated by divergent radiation and determined the maximum size of the pinhole of the camera which is permissible if serious errors are to be avoided. As however the absorption in the specimen has a considerable effect on the breadths of the diffraction lines, certain investigators have modified the formulae and in some cases have used special specimens in endeavouring to circumvent these difficulties. The whole subject has been recently reviewed by Cameron and Patterson (1937) and the relative advantages of each method discussed.

It is the purpose of the present paper to describe an experimental method which enables particle size determinations to be made using a standard cylindrical camera with any reasonable slit system or pinhole, the specimen being mounted in any convenient manner. By mixing a suitable powder consisting of large particles with the powder to be investigated the uncertainties introduced by the experimental conditions can be eliminated.

II. DISCUSSION OF THE NEW EXPERIMENTAL METHOD

(1) *The broadening of the lines due to experimental conditions, assuming monochromatic radiation*

From equation (1) it can be shown that the diffraction broadening is negligible if $t \geqq 10^{-4}$ cm. and we shall give the term "particles of infinite size" to crystals larger than this. The ideal conditions of parallel monochromatic radiation and a point specimen cannot be realized experimentally. Consequently the X-ray lines have a finite breadth, even if the particles composing the specimen are of infinite size. When the particles are in the measurable range the width of the lines is increased beyond the pure diffraction broadening, so that before the theoretical formula can be applied, corrections must be made to the observed breadths. The observed breadths may be influenced by the following factors, other than particle size.

(*a*) The diameter of the cylindrical specimen and the accuracy of centring, if it is rotated.

(*b*) The diameter of the camera, and its distance from the X-ray tube.

(*c*) The absorption of X-rays by the specimen.

(*d*) The pinhole size or slit system of the camera.

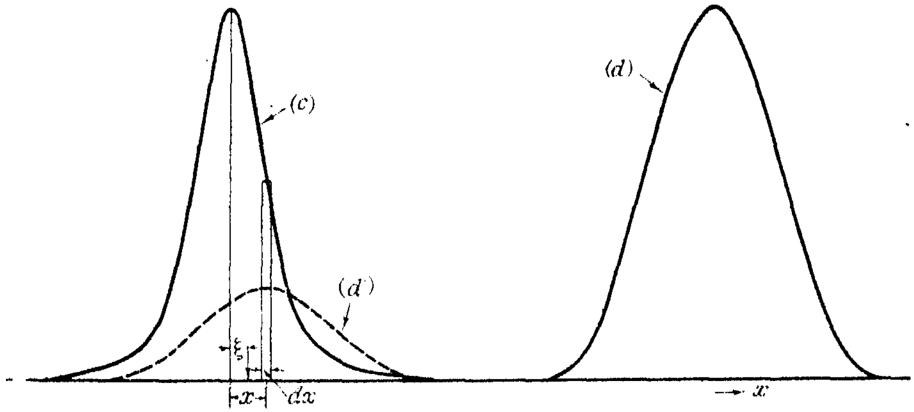
(*e*) The intensity distribution over the target of the X-ray tube considered from the direction in which the X-rays enter the camera. The intensity distribution over the height of the specimen in the beam is governed by (*d*) and (*e*).

The complexity of these factors, some of which are not easy to control, make it very difficult to allow for them by calculation; the advantage of using an experimental method of correcting the breadths is obvious.

The method we have used to correct for this increase in broadening involves the use of a mixed specimen containing not only the particles under investigation but also some particles of infinite size.

The lines produced by such a specimen are of two types: one set produced by the standard substance, called *s*-lines, whose finite breadth is entirely dependent on the experimental conditions; and the other set produced by the substance under investigation—*m*-lines—whose breadth depends on the particle size as well as on the experimental conditions. The breadths of both sets of lines are of course influenced by the angle of reflexion, so that the breadths of the *s*-lines corresponding to the positions of the *m*-lines must be determined by interpolation against the angle of reflexion.

The problem is to determine theoretically the breadth (B) of an *m*-line, given the pure diffraction broadening (β) and the breadth of the *s*-line (b) at the same angle.



$I_s = I_{s(\max)} f(x)$
breadth b

$I_d = I_{d(\max)} F(kx)$
breadth β

FIG. 1

Let us suppose that an *s*-line has the distribution function $I_s(x) = I_{s(\max)} f(x)$ given by fig. 1 (*c*) and hence a breadth b, where

$$b = \frac{\int I_{s(\max)} f(x)\,dx}{I_{s(\max)}} = \int f(x)\,dx.$$

(All breadths are here expressed in linear measure, x being distance along the film from the maximum of the line.)

If the distribution function of the pure diffraction broadening is given by $I_d = I_{d(\max)} F(kx)$ [fig. 1 (*d*)] the breadth is

$$\beta = \int F(kx)\,dx.$$

(The value of $f(x)$ and $F(kx)$ is unity at the maximum in both cases.)

Each element dx of the *s*-line will be broadened, due to diffraction, into a line whose breadth is the same as that given in fig. 1 (*d*) and whose area is equal to that of the element, i.e. $= I_{s(\max)} f(x)\,dx$.

The maximum in the diffracted line corresponding to the element dx will occur at the point x and its magnitude I is given by

$$\frac{I}{I_{d(\max)}} = \frac{I_{s(\max)} f(x)\,dx}{A},$$

where $$A = \int I_{d(\max)}\, F(kx)\,dx.$$

The intensity at any point ξ in the resultant m-line contributed by the element at x is

$$\frac{I_{d(\max)}\, I_{s(\max)}}{A} f(x)\,dx\, F(kx - k\xi).$$

The total intensity at ξ therefore is

$$I_m(\xi) = \frac{I_{d(\max)}\, I_{s(\max)}}{A} \int f(x)\, F(kx - k\xi)\,dx.$$

The maximum intensity of the resultant will occur at $\xi = 0$ and is

$$I_{m(\max)} = \frac{I_{d(\max)}}{A} I_{s(\max)} \int f(x)\, F(kx)\,dx.$$

The total intensity of the resultant line which is obtained by integrating $I_m(\xi)$ over all values of ξ is

$$\int I_m(\xi)\,d\xi = \frac{I_{d(\max)}\, I_{s(\max)}}{A} \iint f(x)\, F(kx - k\xi)\,dx\,d\xi.$$

The breadth of the resultant m-line which is the *observed* breadth is therefore

$$B = \frac{\iint f(x)\, F(kx - k\xi)\,dx\,d\xi}{\int f(x)\, F(kx)\,dx}.$$

Since $$\int_{-\infty}^{+\infty} F(kx - k\xi)\,d\xi = \int F(k\xi)\,d\xi = \int F(kx)\,dx,$$

we can write

$$B = \frac{\int f(x)\,dx \int F(kx)\,dx}{\int f(x)\, F(kx)\,dx}.$$

The true diffraction broadening $\beta = \int F(kx)\,dx$ so that

$$\frac{\text{True diffraction breadth}}{\text{Observed breadth of } m\text{-line*}} = \frac{\beta}{B} = \frac{\int F(kx) f(x)\,dx}{\int f(x)\,dx}. \tag{4}$$

Similarly

$$\frac{\text{Breadth of } s\text{-line}}{\text{Observed breadth of } m\text{-line*}} = \frac{b}{B} = \frac{\int F(kx) f(x)\,dx}{\int F(kx)\,dx}. \tag{5}$$

The parameter k is a function of the particle size. Its variation in (4) and (5) gives different values of b/B and β/B and provided $f(x)$ and $F(kx)$ are known a correction curve can be evaluated giving β/B as a function of b/B.

Evaluation of the correction curve

$f(x)$ was determined experimentally by microphotometering an s-line at $\frac{1}{2}\chi = 80°$, where K_{α_1} and K_{α_2} were well resolved. The resultant curve is shown in fig. 2.

Further measurements showed that providing due allowance was made for the α-doublet the shape of the s-lines was practically independent of the angle χ for specimens which absorb X-rays strongly.†

$F(kx)$ cannot be obtained experimentally in any simple manner. Intensity distributions which have been assumed by Laue are the forms

$$F(kx) = e^{-k^2x^2}, \tag{6}$$

$$F(kx) = \frac{1}{1+k^2x^2}, \tag{7}$$

whilst Bragg's treatment gives

$$F(kx) = \frac{\sin^2 kx}{(kx)^2}. \tag{7a}$$

Using the functions (6) and (7) and the experimental form of $f(x)$ given in fig. 2 the correction curves 3 (a) and 3 (b) were obtained, the integrations being carried out using Simpson's rule. The curve obtained using (7*a*) is almost identical with curve 3 (*a*).

---

* Monochromatic radiation has been assumed. The effect of the α-doublet is considered later.

† This is not the case for a relatively transparent specimen where the intensity distribution at small angles may be complicated by reflexions from both sides of the specimen. Strictly the form of $f(x)$ should be determined for each experimental arrangement.

As a check on the above method we may put $f(x) = \frac{1}{(1+x^2)^2}$, which is a very close approximation to the experimental curve given in fig. 2, then using (7), the correction curve is given by

$$\frac{\beta}{B} = 2p \quad \text{and} \quad \frac{b}{B} = kp,$$

where $p = \frac{2k^3 - 3k^2 + 1}{2(1-k^2)^2}$. This coincides almost exactly with fig. 3 (*b*).

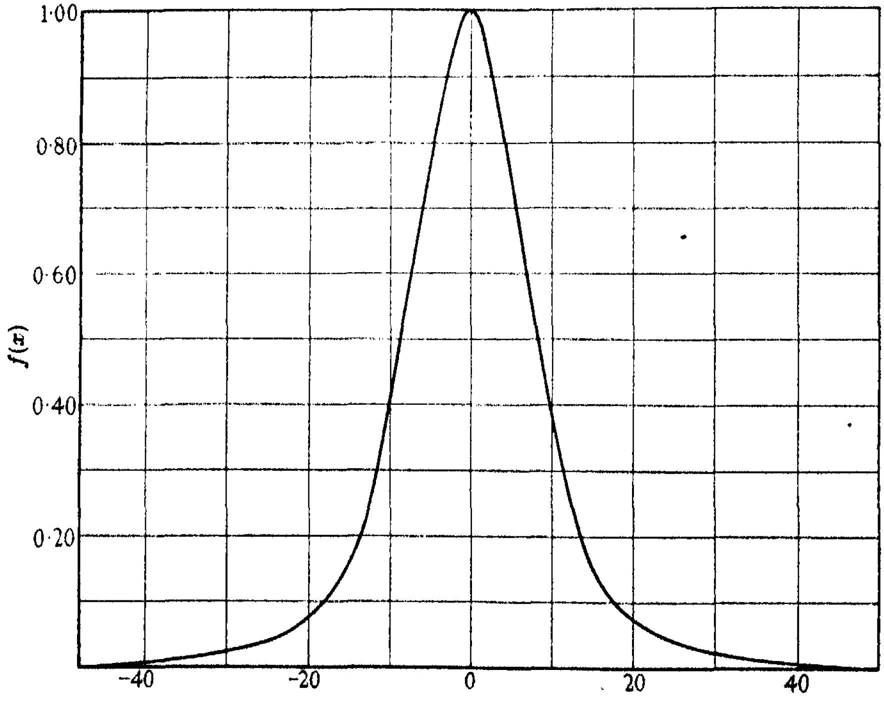


FIG. 2. Line for particles of infinite size, monochromatic radiation.

The difference between curves (*a*) and (*b*) is considerable, especially when the diffraction broadening is small (large particle size). The previous workers, Laue (1926), Brill and Pelzer (1929, 1930), who have attempted to allow for the experimental conditions by calculation, have used only equation (7) for the pure diffraction broadening, so that this ambiguity has not been

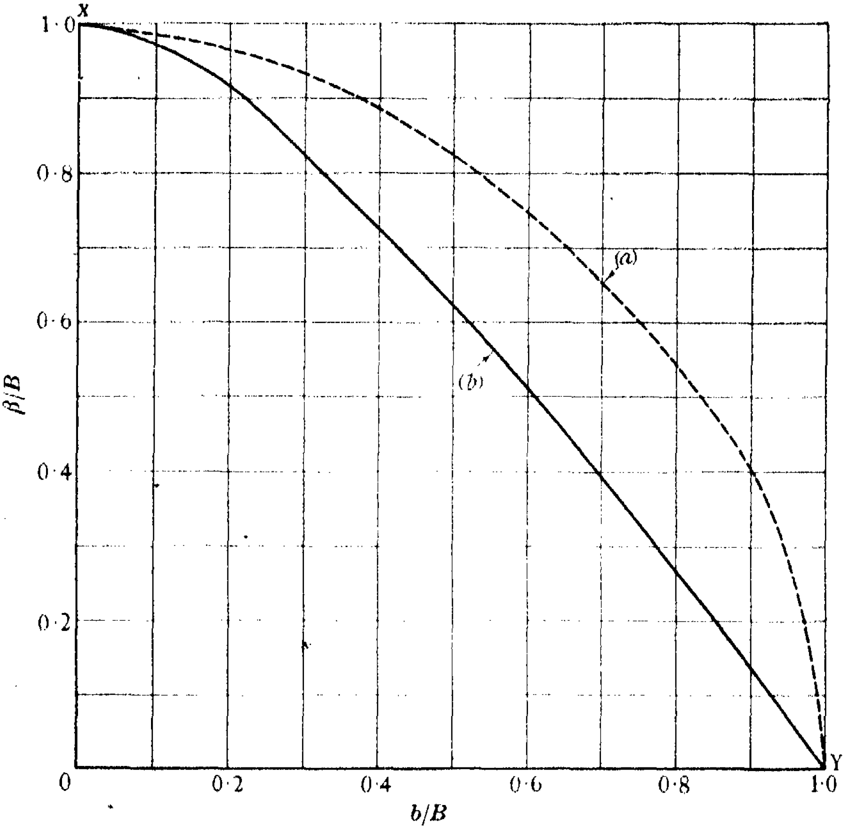


FIG. 3. Correction curve for experimental conditions.

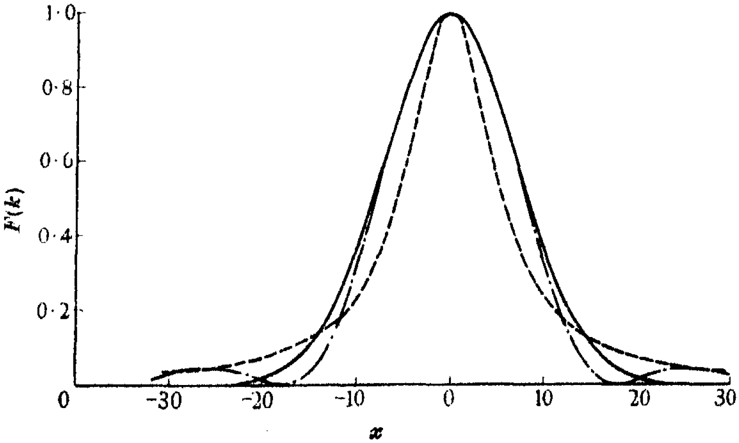


FIG. 4. Line shapes used for $F(kx)$.

noticed previously. As their calculations are in effect the same process as that by which the curve in 3 (*b*) was obtained, there is the possibility of serious error unless the diffraction broadening is of the form given by equation (7).

The curve giving the pure diffraction broadening of the specimen as a whole cannot be found theoretically. Equation (7*a*) is the better distribution if the crystals are all the same size, but this is not to be expected.

If there is a range of particle sizes, equations (6) or (7*a*) will not give the true diffraction broadening of the specimen as a whole. The large particles will sharpen the peak whilst the smaller will broaden the base of the line, which will therefore become more like the shape given by equation (7). (See fig. 4, which shows the lines given by equations (6), (7) and (7*a*), adjusted so that their breadths are the same.)

A test of the correctness of the shape assumed for $F(kx)$ is to build up a line from $f(x)$ and $F(kx)$ of suitable breadths, and compare the result with the shape observed experimentally. This is laborious and gives only a rough guide, but it shows that equation (7) gives the better approximation to the diffraction broadening in the case of the colloidal gold specimen examined.

A further check can be made by finding whether the results remain constant when different parts of the correction curve in fig. 3 are used. This has been done by using different radiations and camera diameters. The results (Table III) using curve (*b*) are reasonably constant for any one line.

Undoubtedly the most satisfactory solution to the problem is to use, as far as possible, the part of the correction curve where b/B is small. Thus if the particle size is large, X-rays of long wave-length should be used and the ratio of the diameter of the camera to that of the specimen should be large.

The correction curve 3 (*b*) differs only slightly from Scherrer's correction formula

$$B = \beta + b,$$

which is represented by a straight line joining the points XY. However, Scherrer's method of finding b is based on the assumption that b is independent of χ, which is incorrect and in certain cases may give negative values for the particle size (Haley and Terrey 1936).

(2) *The effect of the non-homogeneity of the radiation*

Previously we have assumed the radiation to be monochromatic. The K_α radiation generally used consists of a doublet which causes a broadening of the lines, since in general the two components are not resolved. Brill

(1928) has given a correction formula based on the assumption that the lines are triangular in shape. As the lines observed in the present investigation were not triangular and the correction to be applied depends on the shape of the lines, it is desirable that the correction should be worked out using the line shape observed. It is a simple matter to determine the correction curve for any assumed line shape.

We will consider the general case. If $I_{\alpha_2} = \phi(x)$ represents the intensity distribution in the X-ray line produced by the specimen from the K_{α_2} component of the incident X-radiation then the intensity distribution produced from the K_{α_1} component is $2\phi(x-d)$, i.e. a line double the intensity of the K_{α_2} line and shifted from the position of the α_2 by a distance d, the doublet separation.

The breadth (B_0) of the resultant obtained by superimposing the two lines is given :

$$B_0 = \frac{\int I_R\,dx}{I_{R(\max)}}$$

$$= \frac{\int \phi(x)\,dx + 2\int \phi(x-d)\,dx}{I_{R(\max)}}$$

$$= \frac{3\int \phi(x)\,dx}{I_{R(\max)}}.$$

If $\phi(x)$ is known, in general $\int \phi(x)\,dx$ can be obtained using Simpson's rule. The maximum of the resultant $I_{R(\max)}$, and thus the breadth of the resultant, can be found by adding the ordinates of the α_1 and α_2 near their maxima.

Below, the procedure is shown for one value of d, the doublet separation, using the curve given in fig. 2 for $\phi(x)$. For α_2

$$\int I_{\alpha_2}\,dx = 19{\cdot}56,$$

$$\text{Breadth} = \frac{\int I_{\alpha_2}\,dx}{I_{\alpha_2\max}} = \frac{19{\cdot}56}{1}.$$

For the resultant $$\int (I_{\alpha_1} + I_{\alpha_2})\,dx = 58{\cdot}68.$$

The ordinates are given only near the peaks.

x	0	1	2	3	4	5	6	7	8
I_{α_2}	1·000	0·990	0·948	0·890	0·825	0·753	0·680	0·607	0·533
I_{α_1}	1·360	1·506	1·650	1·780	1·896	1·980	2·000	1·980	1·896
I_R	2·360	2·496	2·598	2·670	2·721	2·733	2·680	2·587	2·429

Plotting these shows the maximum to be about 2·738:

$$\text{Resultant breadth} = \frac{58\cdot 68}{2\cdot 738} = 21\cdot 44,$$

$$\frac{B}{B_0} = \frac{\text{True breadth}}{\text{Resultant breadth}} = \frac{19\cdot 56}{21\cdot 44} = 0\cdot 912,$$

$$\frac{d}{B_0} = \frac{\text{Separation}}{\text{Resultant breadth}} = \frac{6}{21\cdot 44} = 0\cdot 280.$$

By repeating for other separations the curve (a) in fig. 5 was obtained, giving B/B_0 as a function of d/B_0.

Fig. 5 (b) gives the correction curve obtained for triangular lines (Brill).

From fig. 5 (a) the observed breadth B_0 can be corrected for the separation of the α-doublet, giving the breadth B.

Although the line shape used for $\phi(x)$ (shown in fig. 2) was that determined for particles of infinite size, it was found to be quite a close approximation when the particles are in the measurable range. As the correction curve does not depend very critically upon the line shape used, the curve in fig. 2 (a) has been used for both s-lines and m-lines.

III. EXPERIMENTAL

(1) *The experimental procedure*

The standard substance is chosen so that few cases occur of the overlapping of lines due to the two substances. Particles having an edge of between 10^{-3} and 10^{-4} cm. are required and these can in general be obtained free from strain by annealing filings of a pure metal just below the sintering temperature. The relative proportions of the two substances are chosen so that they give roughly the same peak intensities. Specimens may be either mounted on a hair, or in tubing, and the photograph taken with a Debye-Scherrer camera having any reasonable slit system or pinhole, the specimen being rotated during the exposure.

After the photograph of the mixed specimen has been taken, the film is microphotometered, densities converted to intensities, and the line breadths found. It is useful to draw a curve giving the linear separation of the doublet d as a function of χ for the radiation and camera used. Then, if b_0 and B_0 are the observed breadths of s-lines and m-lines respectively, the ratios d/b_0 or d/B_0 can be found, and the correction for the α-doublet made to both sets of lines, by multiplying by the appropriate value of (b/b_0) or (B/B_0) as determined from fig 5; b and B are the breadths after correction for the α-doublet.

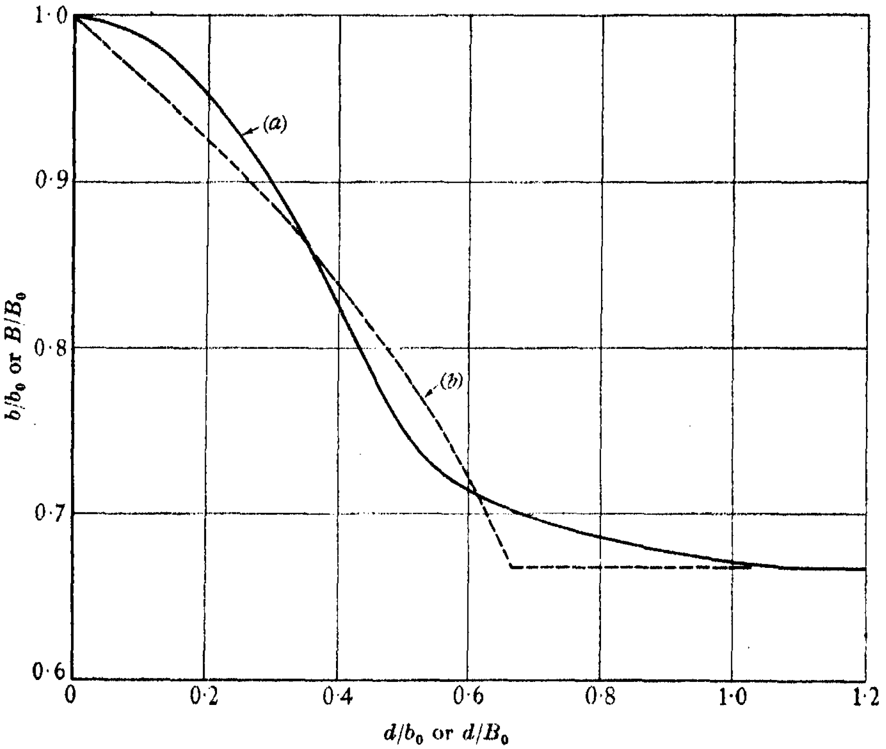


FIG. 5. α-Doublet correction.

The breadths B must now be corrected for experimental conditions. First b is plotted against χ. The ratios b/B can then be found, b being the breadth of the s-line interpolated at the same angle as the m-line. From fig. 3 the corresponding value of (β/B) can be read off. Then $\beta = B(\beta/B)$ is the true diffraction breadth in linear measure.

The angular breadth $\beta_\chi = \beta/R$, where R is the radius of the camera. Then from equation (2) the "apparent particle size" is given in terms of the linear breadth

$$\epsilon = \frac{\lambda R}{\beta \cos \frac{1}{2}\chi_h}. \tag{8}$$

The notation used for the breadths is summarized:

	m-lines	s-lines
(1) Observed breadths	B_0	b_0
(2) After correction for α-doublet	B	b
(3) After correction for experimental conditions	β	—

(2) *Particle size of colloidal gold*

As an example of the use of the method the particle size of a sample of colloidal gold* has been measured, the element molybdenum being used as standard substance.

Copper radiation.	Camera radius, 45·3 mm.
Exposure time, 20 min.	Linear breadths are given in mm.

TABLE I. MOLYBDENUM, s-LINES

Indices	$\theta = \frac{1}{2}\chi$	b_0	d/b_0	b
110	20·3	0·30	0·30	0·27
200	29·5	0·28	0·46	0·22
211	37·0	0·37	0·46	0·29
220	44·0	0·37	0·60	0·27
310	50·9	0·43	0·65	0·30
321	66·1	0·50	1·02	0·33

TABLE II. COLLOIDAL GOLD, m-LINES

Indices	θ	B_0	d/B_0	B	b/B	β	$\cos\theta$	$\beta\cos\theta$	ϵ (A)
111	19·2	0·46	0·17	0·45	0·58	0·24	0·944	0·226	310
200	22·3	0·60	0·15	0·59	0·42	0·42	0·925	0·390	180
220	32·4	0·58	0·24	0·54	0·48	0·35	0·844	0·295	240
311	39·0	0·62	0·29	0·56	0·48	0·36	0·777	0·280	250
222	41·4	0·55	0·35	0·49	0·55	0·28	0·750	0·203	340
331	55·7	0·88	0·37	0·75	0·41	0·54	0·563	0·303	230
422	68·1	1·46	0·38	1·22	0·28	1·02	0·373	0·380	180

---

* The colloidal gold was given to me by Dr J. B. Haley, late of University College, London, to whom I am much indebted.

The values of ϵ should be the same if the particles of colloidal gold were cubic in form. In order to make certain that the variations are due to the crystal form and not the experimental method, the measurements were repeated under other conditions. The results are tabulated below.

TABLE III. VALUES OF ϵ FOR COLLOIDAL GOLD

Radiation	Indices: Camera radius	111	200	220	311	222	331	422
Cu	4·53 cm.	310	180	240	250	340	230	180
Co	4·53 ,,	310	210	230	240	280	170	—
Cu	9·50 ,,	330	190	240	230	330	210	210
Cr	9·50 ,,	300	220	230	260	280		—
Mean (experimental)		310	200	235	245	310	—	—
Calculated		315	212	232	229	315	—	—

The agreement between the experimental results is as good as can be expected. The results obtained by calculation will be referred to in the next section.

IV. DETERMINATION OF THE PARTICLE SHAPE

The method used to determine the particle shape from the different values of the apparent particle size ϵ is essentially that given by Brill (1930). The procedure has however been simplified by the use of "shape curves", which furthermore are of general application.

In the general case of a crystal of parallelepiped form belonging to the monoclinic system

$$\epsilon = \frac{1}{\sqrt{\Sigma\{(\mathbf{b}_i\,\mathbf{G})/m_i\}^2}}, \quad i = 1, 2, 3. \tag{9}$$

The $\mathbf{b}_i$'s are vectors in the reciprocal lattice given by

$$\mathbf{b}_1 = \frac{[\mathbf{a}_2\,\mathbf{a}_3]}{(\mathbf{a}_1\,\mathbf{a}_2\,\mathbf{a}_3)}, \text{ etc.,} \tag{10}$$

where the vectors $\mathbf{a}_i$ define the unit cell. They are chosen so that the volume $(\mathbf{a}_1\,\mathbf{a}_2\,\mathbf{a}_3)$ is positive. **G** is the unit vector in the reciprocal lattice of direction given by

$$\mathbf{G} = \frac{\Sigma h_i \mathbf{b}_i}{|\Sigma h_i \mathbf{b}_i|}, \tag{11}$$

the h_i being the Miller indices of the reflexion. m_i is a pure number, the lengths of the edges of the parallelepiped crystal being $m_1\mathbf{a}_1$, $m_2\mathbf{a}_2$, $m_3\mathbf{a}_3$.

In the simple case of a crystal of the orthorhombic system having three unequal edge lengths it can be easily shown that different values of ϵ are to be expected from different lines.

For crystals of the orthorhombic system equation (9) reduces to

$$\epsilon = \sqrt{\frac{\Sigma(h_i/a_i)^2}{\Sigma(h_i/m_i a_i^2)^2}}. \tag{12}$$

For the reflexion

$$\begin{array}{ll} h_1 00 & \epsilon_{h_1 00} = m_1 a_1, \\ 0h_2 0 & \epsilon_{0h_2 0} = m_2 a_2, \\ 00h_3 & \epsilon_{00h_3} = m_3 a_3. \end{array}$$

Thus, unless the three edge lengths of crystal are all equal, different values of ϵ will be obtained from these three lines, which give the particle shape and size immediately.

Shape determination is more complicated for crystals of the cubic system. All orientations are probable in the specimen, and overlapping occurs, e.g. the 100, 010, 001 reflexions overlap to form one line.

Suppose one line of breadth β is composed of n overlapping reflexions of breadths $\beta_1, \beta_2, \ldots \beta_i, \ldots \beta_n$ corresponding to different ϵ values $\epsilon_1, \epsilon_2, \ldots \epsilon_i, \ldots \epsilon_n$. If the total intensity $\int I_i \, d\chi$ is the same for each reflexion

$$\beta = \frac{n\int I_i \, d\chi}{\Sigma I_{i\,(\max)}}.$$

Since

$$I_{i\,(\max)} = \frac{\int I_i \, d\chi}{\beta_i},$$

$$\beta = \frac{n}{\Sigma 1/\beta_i}.$$

Now

$$\beta = \frac{\lambda}{\epsilon \cos \frac{1}{2}\chi}.$$

$\therefore$

$$\epsilon = \frac{1}{n}\Sigma\epsilon_i. \tag{13}$$

For crystals of the cubic system equation (12) reduces to

$$\epsilon = a\sqrt{\frac{h_1^2+h_2^2+h_3^2}{(h_1/m_1)^2+(h_2/m_2)^2+(h_3/m_3)^2}}. \tag{14}$$

Simultaneous variation of m_1, m_2, m_3 is too complicated. Brill takes $m = m_1 = m_2 = pm_3$. Variation of the parameter p covers the shapes of both a needle ($p < 1$) and a plate ($p > 1$).

Then
$$\frac{\epsilon}{ma} = \sqrt{\frac{h_1^2+h_2^2+h_3^2}{h_1^2+h_2^2+p^2h_3^2}}. \tag{15}$$

In the following $\epsilon_{|111|}$ denotes the ϵ values obtained from the 111 line due to the eight overlapping reflexions. Similarly for other indices.

From equations (13) and (15)

$$\frac{\epsilon_{|111|}}{ma} = \sqrt{\frac{3}{2+p^2}},$$

$$\frac{\epsilon_{|200|}}{ma} = \frac{1}{3}\left(1+1+\frac{1}{p}\right),$$

$$\frac{\epsilon_{|220|}}{ma} = \frac{1}{3}\left(1+2\sqrt{\frac{2}{1+p^2}}\right),$$

$$\frac{\epsilon_{|311|}}{ma} = \frac{1}{3}\left(2\sqrt{\frac{11}{10+p^2}}+\sqrt{\frac{11}{2+9p^2}}\right).$$

$\epsilon_{|222|}$ is the same as $\epsilon_{|111|}$.

Fig. 6 shows $\epsilon_{|h_1h_2h_3|}/ma$ plotted against $\log_{10} p$, using the equations given above. No value of p gives the same relative ϵ values as those obtained with the colloidal gold. It is noticeable that high ϵ values are obtained when p is small and the reflecting planes are parallel to the faces of the unit cell. As we require $\epsilon_{|111|}$ to be high, a change to a unit cell with its base on the cubic (111) plane is indicated. The lattice parameters of the new orthohexagonal cell are

$$a_1 = \sqrt{3}\,a; \quad a_2 = \sqrt{\tfrac{3}{2}}\,a; \quad a_3 = \sqrt{\tfrac{1}{2}}\,a,$$

in the directions perpendicular to the cubic planes (111), $(2\bar{1}\bar{1})$, $(01\bar{1})$ respectively.

The new indices of the various planes which were found by plotting the reciprocal lattice are as follows:

Cubic	Orthohexagonal
8 × 111	2 × 120; 4 × 111; 2 × 300
6 × 200	2 × 220; 4 × 211
12 × 220	4 × 411; 2 × 420; 4 × 031; 2 × 002
24 × 311	2 × 520; 4 × 511; 8 × 331; 2 × 140; 4 × 122; 4 × 302

For an orthorhombic lattice

$$\epsilon = \sqrt{\frac{\Sigma(h_i/a_i)^2}{\Sigma(h_i/m_i a_i^2)^2}}. \tag{16}$$

Putting in the values for a_1, a_2, a_3 gives

$$\epsilon = \sqrt{\frac{h_1^2/3a^2 + 2h_2^2/3a^2 + 2h_3^2/a^2}{h_1^2/9m_1^2a^4 + 4h_2^2/9m_2^2a^4 + 4h_3^2/m_3^2a^4}}.$$

Taking

$$m = \sqrt{\tfrac{3}{2}}\,m_2 = \sqrt{\tfrac{1}{2}}\,m_3 = \sqrt{3}\,qm_1,$$

$$\epsilon = ma\sqrt{\frac{h_1^2 + 2h_2^2 + 6h_3^2}{q^2h_1^2 + 2h_2^2 + 6h_3^2}}. \tag{17}$$

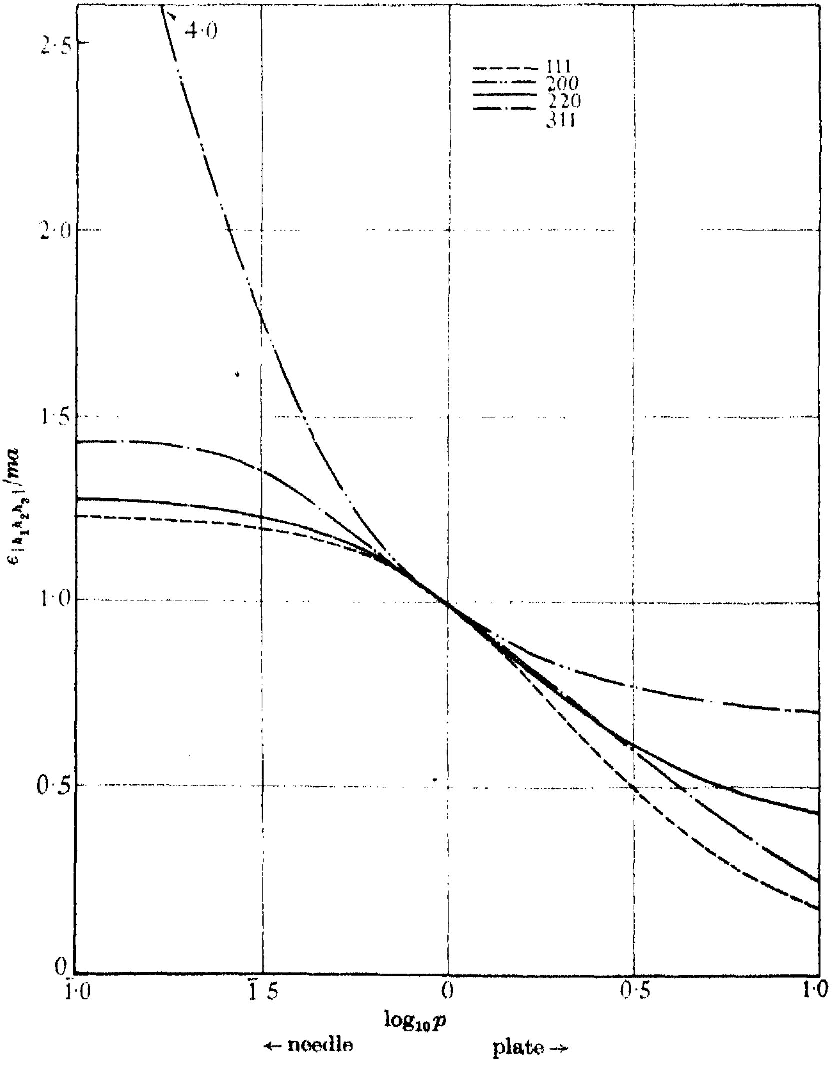


FIG. 6. Shape curve; cubic unit cell.

To take one example, and still designating ϵ by the cubic indices, we get from equations (13) and (17)

$$\frac{\epsilon_{|220|}}{ma}=\frac{1}{12}\left\{4\sqrt{\frac{16+8}{16q^2+8}}+2\sqrt{\frac{16+8}{16q^2+8}}+4+2\right\}$$
$$=\frac{1}{2}\left\{\sqrt{\frac{3}{2q^2+1}}+1\right\}.$$

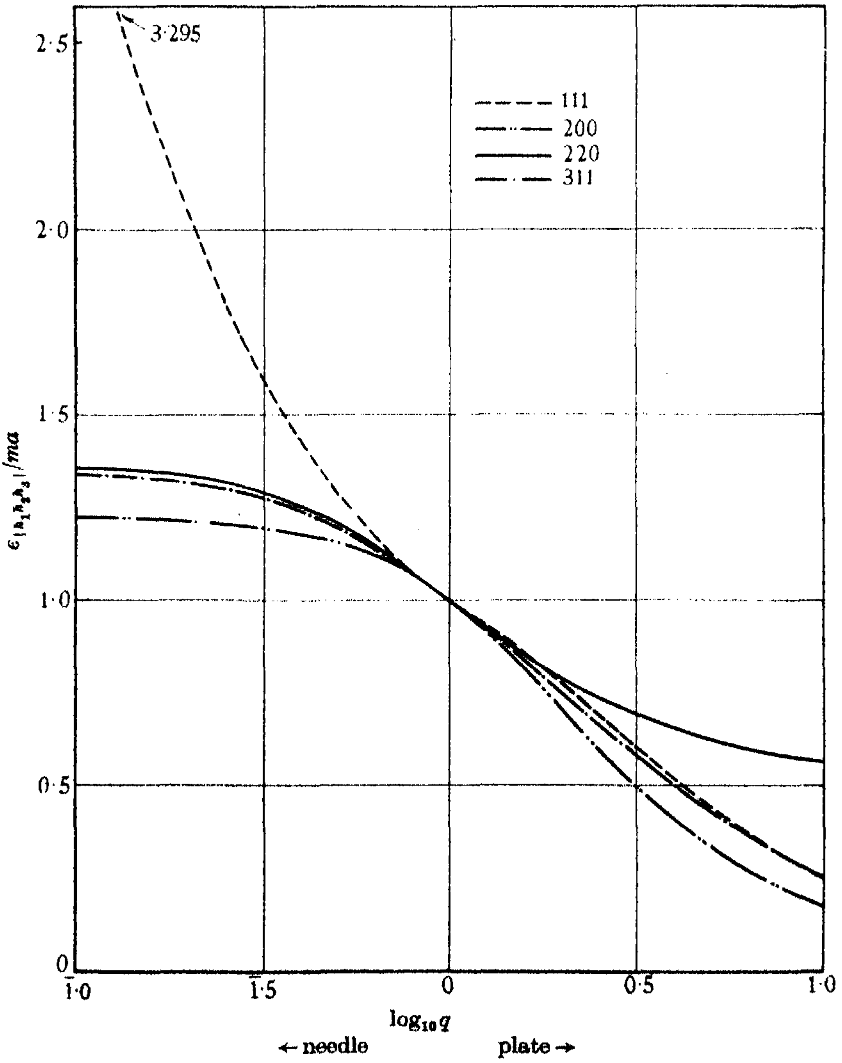


FIG. 7. Shape curve; orthogonal unit cell.

In fig. 7 $\epsilon_{|h_1h_2h_3|}/ma$ is plotted against $\log_{10} q$. It can be seen that $\log_{10} q = \bar{1}\cdot 4$ ($q = 0\cdot 25$) gives roughly the observed ratios of the ϵ values.

From the curve, by addition

$$(\epsilon_{|111|} + \epsilon_{|200|} + \epsilon_{|220|} + \epsilon_{|311|} + \epsilon_{|222|}) = 7\cdot 41\ ma.$$

From Table III,

$$1300 = 7\cdot 41\ ma,$$

$$ma = 176\ \text{A}.$$

The dimensions of the crystal are given:

$$m_1 a_1 = \frac{m}{\sqrt{3 \times 0\cdot 25}} \times \sqrt{3}\, a = 703\ \text{A},$$

$$m_2 a_2 = \sqrt{\tfrac{2}{3}}\, m \times \sqrt{\tfrac{3}{2}}\, a = 176\ \text{A},$$

$$m_3 a_3 = \sqrt{2}\, m \times \sqrt{\tfrac{1}{2}}\, a = 176\ \text{A}.$$

The calculated ϵ values using these dimensions are those given in Table III.

Although the agreement with the observed ϵ values is good, the calculated shape is not very convincing. We restrict ourselves to pointing out the agreement, rather than asserting that the shape and orientation given above are correct.

V. THEORETICAL

(1) *Laue's treatment*

We shall consider Part II, § 4 of his paper. Two equations are derived, either of which gives the intensity distribution in the X-ray diffraction line as a function of the particle size. In the course of the derivation the "interference function" $f(A_i)$* must be integrated.

According to Laue

$$f_0(A_i) = \frac{\sin^2 M_1 A_1}{\sin^2 \frac{1}{2} A_1} \frac{\sin^2 M_2 A_2}{\sin^2 \frac{1}{2} A_2} \frac{\sin^2 M_3 A_3}{\sin^2 \frac{1}{2} A_3} \tag{18}$$

for a parallelepiped of edges $M_1 \mathbf{a}_1$, $M_2 \mathbf{a}_2$, $M_3 \mathbf{a}_3$, where the vectors $\mathbf{a}_i$ define the unit cell.

Actually, as Laue states in another publication (Laue 1915), this equation refers to a parallelepiped of edges $2M_1 \mathbf{a}_1$, $2M_2 \mathbf{a}_2$, $2M_3 \mathbf{a}_3$. The correct interference function for a parallelepiped of edges $M_i \mathbf{a}_i$ is

$$f_0(A_i) = \frac{\sin^2 \frac{1}{2} M_1 A_1}{\sin^2 \frac{1}{2} A_1} \frac{\sin^2 \frac{1}{2} M_2 A_2}{\sin^2 \frac{1}{2} A_2} \frac{\sin^2 \frac{1}{2} M_3 A_3}{\sin^2 \frac{1}{2} A_3}. \tag{19}$$

---

* See Laue's paper for the meaning of A_i. His notation will be used throughout.

This arithmetical mistake leads to the particle size being in error by a factor of 2 and will be referred to later.

To reduce integration difficulties, Laue replaces equation (18) by a function of the form

$$f(A_i) = \phi\{\Sigma m_i^2(A_i - 2\pi h_i)^2\},$$

where the h_i are the Miller indices of the reflexion.

Two examples of such a function are

$$f_1(A_i) = \rho_1 e^{-\omega_1^2 \Sigma m_i^2(A_i - 2\pi h_i)^2}, \tag{20}$$

$$f_2(A_i) = \frac{\rho_2}{\{\Sigma m_i^2(A_i - 2\pi h_i)^2 + \omega_2^{-2}\}^2}. \tag{21}$$

Laue chooses the values of the parameters ρ and ω, so that if $M_i = m_i$ these equations give the same value for the maximum and the integral value over the volume surrounding the whole number point $A_i = 2\pi h_i$ as equation (18). In particular he finds

$$\omega_1 = \frac{1}{\sqrt{\pi}} = \frac{1}{1{\cdot}77},$$

$$\omega_2 = \frac{1}{\sqrt[3]{2\pi}} = \frac{1}{1{\cdot}85}.$$

The intensity distribution in the X-ray line, derived from (20) and (21), is given

from (20) $$J_1(\chi) = \text{const.}\, e^{-\frac{\omega_1^2}{\eta^2}\delta^2}, \tag{22}$$

from (21) $$J_2(\chi) = \frac{\text{const.}}{\frac{\omega_2^2}{\eta^2}\delta^2 + 1}, \tag{23}$$

where $\frac{1}{2}\chi = \theta$, the Bragg angle,

$\delta = \sin\frac{1}{2}\chi - \sin\frac{1}{2}\chi_h$, χ_h being the position of the maximum.

η is a pure number which depends only on the size and shape of the crystal and the wave-length of the X-radiation used.

$$\eta = \frac{1}{2k}\sqrt{\Sigma\left\{\frac{(\mathbf{b}_i \mathbf{G})}{m_i}\right\}^2}, \tag{24}$$

$$k = \frac{2\pi}{\lambda}. \tag{25}$$

$\mathbf{b}_i$, $\mathbf{G}$ and m_i have been defined in equations (10) and (11). It should be noted that for convenience we have used $\epsilon = \frac{1}{2k\eta}$. Reverting to equations (22) and (23) the line breadths β can be calculated from the definition

$$\beta = \frac{\int_{-\infty}^{+\infty} J(\chi)\,d\chi}{J(\chi)_{\max}}.$$

In the scale of δ (instead of χ) these are

from (20) $$\beta_1 = \frac{\eta\sqrt{\pi}}{\omega_1}, \tag{26}$$

from (21) $$\beta_2 = \frac{\eta\pi}{\omega_2}. \tag{27}$$

Substituting $\omega_1 = \frac{1}{\sqrt{\pi}}$, $\omega_2 = \frac{1}{\sqrt[3]{2\pi}}$ and converting from the scale of δ to the scale of χ by multiplication by $\frac{2}{\cos\frac{1}{2}\chi_h}$ we obtain

$$\beta_1 = \frac{2\eta\pi}{\cos\frac{1}{2}\chi_h}, \quad \beta_2 = \frac{3\cdot 7\eta\pi}{\cos\frac{1}{2}\chi_h}.$$

For particles of the cubic system where

$$|m_1\mathbf{a}_1| = |m_2\mathbf{a}_2| = |m_3\mathbf{a}_3| = t,$$

$$\eta = \frac{\lambda}{4\pi t},$$

so that $$\beta_1 = 0\cdot 5\frac{\lambda}{t\cos\frac{1}{2}\chi_h}, \tag{26a}$$

$$\beta_2 = 0\cdot 92\frac{\lambda}{t\cos\frac{1}{2}\chi_h}. \tag{27a}$$

Laue uses equation (23) as the intensity distribution and shows that (27*a*) is in reasonable agreement with the formula of Scherrer in which

$$\beta = \frac{0\cdot 94\lambda}{t\cos\frac{1}{2}\chi_h}.$$

Actually this agreement is purely fortuitous since both (26*a*) and (27*a*) contain an arithmetical error of 2 due to equation (18) being used instead of equation (19), and furthermore equation (27*a*) contains an additional error introduced by an arithmetical mistake in the evaluation of ω_2. According to Laue $\omega_2 = \frac{1}{\sqrt[3]{2\pi}}$, whereas the correct value, determined by his method, is $\omega_2 = \frac{1}{\sqrt[3]{\pi}}$.

If these corrections are made

$$\beta_1 = \frac{\lambda}{t \cos \frac{1}{2}\chi_h}, \tag{26b}$$

$$\beta_2 = \frac{1{\cdot}46\lambda}{t \cos \frac{1}{2}\chi_h}. \tag{27b}$$

(26*b*) is in much better agreement with Scherrer's formula than (27*b*) but the agreement between the equations (26*b*) and (27*b*) is not convincing.

(2) *Alternative method for evaluation of the constant* ω

The lack of agreement between equations (26*b*) and (27*b*) suggested that the method used by Laue to fit $f_1(A_i)$ and $f_2(A_i)$ to the interference function $f_0(A_i)$ might be improved and led to the following treatment.

By way of analogy, fig. 8 (*a*) shows a whole number point S (where $A_i = 2\pi h_i$) in an imaginary reciprocal lattice of only two dimensions, O being the origin and $A_i\mathbf{b}_i$ the axes. The value of the interference function $f_1(A_i)$ which is a maximum at S, and for a powder specimen of crystal grains of different sizes and orientations may be considered to decrease in all directions proceeding from S, is represented by contour lines around S. Laue has shown that the intensity in the X-ray line at an angle χ is proportional to the integral of $f(A_i)$ taken over a line PQ perpendicular to, and passing through the end point of the vector $\mathbf{H}$ in the direction OS, where $|\mathbf{H}| = 2 \sin \frac{1}{2}\chi$. The integral of $f(A_i)$ over PQ is a maximum when PQ passes through S.

The breadth β of the X-ray line is proportional to $\int I/I_{\max}$, where $\int I$ is the integral of $f(A_i)$ over the volume surrounding S, i.e.

$$\iiint f(A_i)\, dA_1 dA_2 dA_3,$$

and $I_{\max}$ is the integral of $f(A_1)$ over the plane PQ through S. We shall fit $f_1(A_i)$ in equation (20) and $f_2(A_i)$ in equation (21) to $f_0(A_i)$ by choosing the values of ω_1 and ω_2 so that the ratio $\int I/I_{\max}$ is the same for all three functions. The correct value of $f_0(A_i)$, i.e. equation (19), is used and the results are given in the scale of A_1. The integrations have been done for the $(h_1 00)$ reflexion of an orthorhombic lattice (fig. 8 (*b*)), as this is reasonably straightforward. The value of ω should be independent of the particular reflexion and lattice considered:

(i) $$f_0(A_i) = \frac{\sin^2 \frac{1}{2}M_1 A_1}{\sin^2 \frac{1}{2}A_1} \cdot \frac{\sin^2 \frac{1}{2}M_2 A_2}{\sin^2 \frac{1}{2}A_2} \cdot \frac{\sin^2 \frac{1}{2}M_3 A_3}{\sin^2 \frac{1}{2}A_3}.$$

Using the integration limits

$$A_1 = 2\pi h_1 \pm \tfrac{1}{2}\pi, \quad A_2 = A_3 = \pm \tfrac{1}{2}\pi,$$

$$\int I = 8 M_1 M_2 M_3 \pi^3,$$

$$I_{\max} = 4 M_1^2 M_2 M_3 \pi^2,$$

$$\beta_0' = \frac{2\pi}{M_1}.$$

(ii) $$f_1(A_i) = \rho_1 e^{-\omega_1^2 \Sigma m_i^2 (A_i - 2\pi h_i)^2}.$$

Using the integration limits ± ∞

$$\int I = \frac{\rho_1 \pi^{\frac{3}{2}}}{\omega_1^3 m_1 m_2 m_3},$$

$$I_{\max} = \frac{\rho_1 \pi}{\omega_1^2 m_2 m_3},$$

$$\beta_1' = \frac{\sqrt{\pi}}{\omega_1 m_1}.$$

(iii) $$f_2(A_i) = \frac{\rho_2}{\{\Sigma m_i^2 (A_i - 2\pi h_i)^2 + \omega_2^{-2}\}^2}.$$

Using the integration limits ± ∞

$$\int I = \frac{\omega_2 \pi^2 \rho_2}{m_1 m_2 m_3},$$

$$I_{\max} = \frac{\pi \rho_2 \omega_2^2}{m_2 m_3},$$

$$\beta_2' = \frac{\pi}{\omega_2 m_1}.$$

The integration limits ± ∞ can be used for $f_1(A_i)$ and $f_2(A_i)$ as these functions are not periodic in A_i.

If $M_i = m_i$ and the values $\beta'_0, \beta'_1, \beta'_2$ are to agree

$$\omega_1 = \frac{1}{2\sqrt{\pi}} \tag{28}$$

$$\omega_2 = \tfrac{1}{2}. \tag{29}$$

If these values of ω are used the same value for the particle size is obtained from both equations (26) and (27), viz.

$$\beta = 2\pi\eta.$$

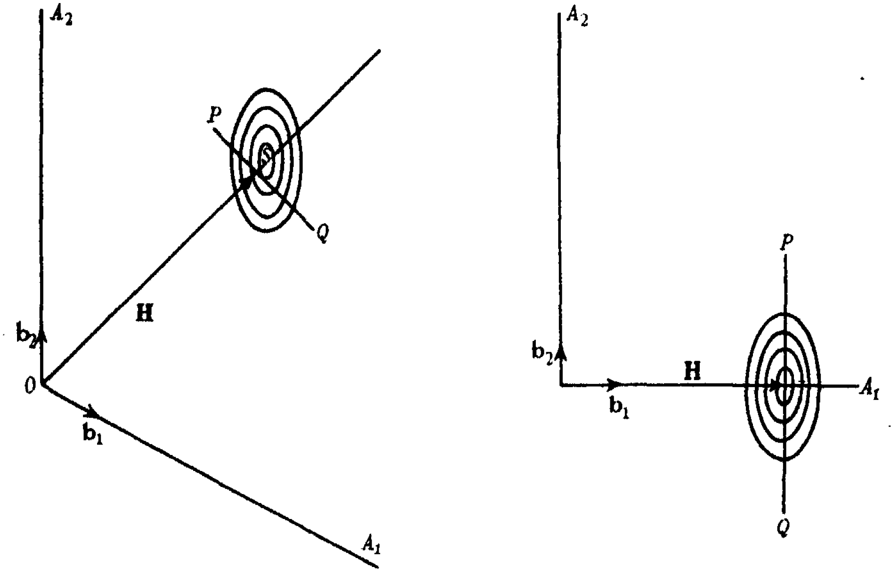


Fig. 8

Reverting to the scale of χ,

$$\beta = \frac{4\pi\eta}{\cos\frac{1}{2}\chi_h}.$$

Substituting $\eta = \dfrac{\lambda}{4\pi t}$ for comparison with Scherrer's formula,

$$\beta = \frac{\lambda}{t\cos\frac{1}{2}\chi_h},$$

which is in reasonable agreement.

When the particles are of general parallelepiped form the line breadths are given by

$$\beta_\chi = \frac{\lambda}{\epsilon\cos\frac{1}{2}\chi_h}, \tag{30}$$

where $$\epsilon = \frac{1}{2k\eta} = \frac{1}{\sqrt{\Sigma\{(\mathbf{b}_i\,\mathbf{G})/m_i\}^2}}. \tag{31}$$

In equation (2) the breadth β_χ has been given the suffix χ merely to distinguish it from the linear breadth which has been denoted as β in the experimental section.

VI. THE EFFECT OF THE PARTICLE SIZE DISTRIBUTION

It is not to be expected that all the particles composing the specimen are the same size and shape. Consequently the observed mean size will differ from the true mean size.

Patterson (1928) has attempted to find the effect of a Maxwellian distribution. He compared Laue's result, which uses equation (23) for the intensity distribution in the X-ray line and assumes all particles are the same size, with the result obtained using equation (22) and assuming a Maxwellian distribution of the particles. He commences with equation (22),

$$J(\chi) = \text{constant}\, e^{-\frac{\omega^2}{\eta^2}(\sin\frac{1}{2}\chi - \sin\frac{1}{2}\chi_h)^2},$$

but has ignored the fact that the "constant" is a constant for one size of particle only. Assuming cubic crystals and $m_1 = m_2 = m_3 = m$, this equation should be

$$J_{(\chi)} = \text{constant}'\, \eta^{-4} e^{-\frac{\omega^2}{\eta^2}(\sin\frac{1}{2}\chi - \sin\frac{1}{2}\chi_h)^2},$$

by reasoning similar to that used in the derivation of equation (32). This omission invalidates the calculation.

His result, that the observed mean value of η differs from the true mean value of η by $\sqrt{\pi}$, follows directly from the incorrect assumption that ω_1 and ω_2 in equations (22) and (23) are the same. We have shown that these constants differ by $\sqrt{\pi}$.

The law governing the particle size distribution is not known and it is highly probable that it will vary from powder to powder, depending to a great extent on the method of production of the powder. Nevertheless some estimate of the effect of the variation in particle size is desirable. A number of assumptions for the distribution law will be made and the ratio between the true mean size and the size as determined by X-ray measurements calculated for the different assumptions.

We will consider the case for cubic crystals where $m_1 a = m_2 a = m_3 a = t$.

The total intensity reflected by the crystal is proportional to its volume, i.e.

$$\int_{-\infty}^{+\infty} J_{(\chi)}\, d\chi = c_1 t^3.$$

From equation (1) the breadth

$$\beta = \frac{c_2}{t} = \frac{\int J_{(\chi)}\,d\chi}{J_{\max}} = \frac{c_1 t^3}{J_{\max}}.$$

$$\therefore \qquad J_{\max} = \frac{c_1}{c_2} t^4.$$

c_1 and c_2 are constants for one line, which are independent of the particle size.

If there are n particles of size t, the breadth of the line due to all the particles

$$\beta_{\Sigma n} = c_2 \frac{\Sigma n t^3}{\Sigma n t^4}.$$

The observed mean size is therefore

$$t_{\text{obs. mean}} = \frac{\Sigma n t^4}{\Sigma n t^3}, \tag{32}$$

compared with

$$t_{\text{true mean}} = \frac{\Sigma n t}{\Sigma n}. \tag{33}$$

Let

$$\gamma = \frac{t_{\text{true mean}}}{t_{\text{obs. mean}}}. \tag{34}$$

Then γ can be found from equations (32) and (33) if the distribution law is known.

Some form of skew distribution law is most probable, the objection to the Gaussian law being that particles cannot be negative in size.

However, as the modulus of distribution is unknown we cannot hope to get more than an idea of the order of γ. We have therefore assumed for the sake of simplicity that the distribution is Gaussian. Equation (32) shows that the larger particles are given more weight, so that if h, the modulus of distribution, is adjusted so that the number of particles of negative size is small, the approximation will be sufficiently close for our purpose.

The number of particles of size t is

$$dn = A e^{-h^2(t-t_0)^2}\,dt,$$

where t_0 is the true mean size, and A a constant. Then

$$\gamma = t_0 \frac{\int_{-\infty}^{+\infty} t^3 e^{-h^2(t-t_0)^2}\,dt}{\int_{-\infty}^{+\infty} t^4 e^{-h^2(t-t_0)^2}\,dt},$$

i.e.

$$\gamma = \frac{t_0^4 + \dfrac{3t_0^2}{2h^2}}{t_0^4 + \dfrac{3t_0^2}{h^2} + \dfrac{3}{4h^4}}. \tag{35}$$

As an example, take $t_0 = 100\,\text{A}$, $h = \frac{2}{100}$, which gives the number of particles for $t = 0$ less than 2 % of that for $t = t_0$. Then $\gamma = \frac{88}{115} = 0{\cdot}765$.

If all the particles are the same size $h = \infty$ and $\gamma = 1$. As h decreases the spread of the distribution increases and γ decreases. But if h is smaller than $2/t_0$ (the value taken in the example) a considerable number of particles will have negative sizes. Thus it is unlikely that γ will be much smaller than about 0·75.

If the distribution were Maxwellian we might assume, following Patterson,

$$dn = A_1 t^2 e^{-k_1 t^2}\,dt,$$

which gives a value for γ of 0·68.

Alternatively to the above, the assumption might be made that the distribution law applies to the volumes, instead of the edges of the particles.

In spite of the many uncertainties the conclusion can be drawn that the observed mean size will be greater than the true mean size, and it is our opinion that the difference between them is unlikely to be greater than 30 % of the observed mean size.

ACKNOWLEDGEMENT

I am indebted to Dr C. Sykes for suggesting the problem and also for his advice throughout the investigation.

My thanks are due to the Metropolitan-Vickers Electrical Co. Ltd. for financial assistance and also to Dr A. P. M. Fleming, C.B.E., Director and Manager of Research and Education Departments, Metropolitan-Vickers Electrical Co. Ltd., for his personal interest in the work.

Most of the experimental work was done in the Physical Laboratories of the University of Manchester and my thanks are due to Professor W. L. Bragg for his kind interest and supervision.

REFERENCES

Bragg 1933 "Crystalline State", **1**, 189.
Brill 1928 *Z. Krystallogr.* **68**, 387.
— 1930 *Z. Krystallogr.* **75**, 217.
Brill and Pelzer 1929 *Z. Krystallogr.* **72**, 398.
— — 1930 *Z. Krystallogr.* **74**, 147.

Cameron and Patterson 1937 American Society for Testing Materials. "Symposium on Radiography and X-ray Diffraction, 1937."
Haley and Torrey 1936 *Trans. Faraday Soc.* **32**, 1313.
Laue 1915 *Enzykl. math. Wiss.* Band 5, Teil 3, Wellenoptik, p. 460.
— 1926 *Z. Krystallogr.* **64**, 115.
Patterson 1928 *Z. Krystallogr.* **66**, 637.
Scherrer 1920 "Zsigmondy Kolloidchemie", 3rd edn. p. 387.
Seljakow 1925 *Z. Phys.* **31**, 439.

The resistance of superconducting cylinders in a transverse magnetic field

By A. D. Misener, *Emmanuel College, Cambridge*

(*Communicated by J. D. Cockcroft, F.R.S.—Received* 29 *January* 1938)

1. Introduction

Experiments on the penetration of a magnetic field (de Haas and Casimir-Jonker 1934) into superconductors have shown that, when a superconducting cylinder is placed in an increasing transverse field, penetration of the field first occurs when the applied field strength reaches a value $0{\cdot}50H_k$, where H_k is the critical field corresponding to the temperature of the experiment. Since, for this value of the applied field, the field strength, at the surface of the cylinder (von Laue 1932) where it is intersected by a diametral plane perpendicular to the direction of the field, will be precisely H_k, the above result is in accordance with expectation.

On the other hand, it was found by de Haas, Voogd and Jonker (1934) that under the same conditions the cylinder first exhibited electrical resistance when the applied field strength reached the value $0{\cdot}58H_k$. Since this discrepancy probably results from the properties of the "intermediate state" (Peierls 1936; London 1936; Landau 1937) occurring when the magnetic field just begins to penetrate the superconductor, it seemed desirable to investigate the matter in more detail.

The main difficulty in the experiment arises in the determination of the exact field strength (H_t) at which the first traces of resistance appear. The resistance at first increases very gradually with increasing magnetic field, so H_t cannot be determined with sufficient accuracy by extrapolating the transition curves unless these are determined very precisely.

The experiments described below were designed to provide the required accuracy particularly near H_t where the resistance to be measured is very small. Cylindrical polycrystalline wires were used, the best results being obtained with tin and indium. It was found that the ratio H_t/H_k was not a constant but varied linearly with temperature.

2. Experimental method

The potential difference between two points on the wire was observed with a constant current flowing in the wire and at a constant temperature below the normal transition point ($H = 0$). The readings were invariably taken with the external field increasing by steps of a few gauss from zero.

(*a*) The wires used were made by extruding, at a temperature 20° C. below the melting-point, a quantity of the pure metal through steel dies. The usual rate was 3–4 cm./min. The cross-section was circular to within 1 % and of diameter 0·010 or 0·015 cm. The crystallites were small and showed no preferred orientation. Specimens of 6–7 cm. length were cut from these wires; they were mounted straight and vertically in a liquid helium cryostat and were in contact with the liquid helium everywhere between the two points selected for the potential measurements. At these points two short leads of the same material as the specimen were fused to it, and to the ends of these were soldered leads of 38-gauge manganin wire. Similar current leads were soldered to the ends of the specimen.

(*b*) The purity of the individual specimens was tested, after mounting, by determining the resistance ratio R/R_0 at 4·2° K. (R_0 = room temperature resistance) and by the abruptness of the transition curve in a magnetic field applied parallel to the axis of the wire. All specimens used had resistance ratios between 10^{-4} and 7×10^{-4}.

(*c*) The resistance measurements were made with a potentiometer bridge similar in arrangement to that described by Kapitza and Milner (1937). The sensitivity was 10^{-10} V/mm. The usual measuring current in the specimen was 10 mA,* which was sufficient to detect $10^{-4} R_n$ (R_n = resistance in the normal state just above the transition point). From graphs of R against H the values H_t ($R \rightarrow 0$) were determined.

---

* Determinations of the effect of current strength on the shape of the transition curve, to be published shortly, show that for all reasonably small currents the point at which resistance first appears is independent of current strength.

(*d*) The transverse magnetic field was produced by an electromagnet with large flat pole-pieces which gave a field uniform to within 1 % over a diameter of 7 cm. The longitudinal field was obtained from an air core solenoid. All fields were calibrated against standard coils using a fluxmeter. Temperatures were calculated from the vapour pressure of the liquid helium bath surrounding the specimen. By observing a butyl-phthalate pressure difference gauge the temperature could be kept constant to within 0·01° K.

(*e*) The values of H_k appropriate to the temperatures at which the different H_t's were determined could be obtained from the published results on critical fields (Schubnikow and others 1936; de Haas and Engelkes 1937; Daunt and Mendelssohn 1937). The critical field (longitudinal) was also determined for each specimen at several temperatures. The values agreed with those quoted in the above references to within 1 %. It was found, however, that equally accurate values of the critical field could be obtained from the transition curve in a transverse magnetic field. If the linear portion of this curve is extrapolated upwards to meet the extension of the normal resistance curve at the point A (fig. 1) it was found that the value of the

TABLE I

Temp. ° K.	H_k	H_k (*a*)	H_k (*b*)
3·33	54·5	54·0	54
3·12	81	82·0	81
2·92	109	109	108
2·71	138	135	134
2·29	182	182	179
1·49	248	250	248

(*a*) de Haas and Engelkes. (*b*) Daunt and Mendelssohn.

TABLE II

Temp.	H_k		Temp.	H_k	
° K.	Long.	Trans.	° K.	Long.	Trans.
3·31	10	10	2·27	150	—
3·15	33	33	2·24	152	—
3·12	39	—	2·17	—	160
3·04	49	50	2·07	—	170
3·03	51	—	1·68	—	204
2·98	58	58	1·61	209	—
2·95	60	61	1·58	—	214
2·75	89	—	1·56	215	—
2·67	—	100	1·50	219	221
2·53	121	—	1·38	227	—
2·46	—	125			

field corresponding to A was precisely H_k. Values for the critical field of tin obtained in this manner are compared, in Table I, with the published data. No very extensive data on the critical field of indium could be found, so H_k was determined in both longitudinal and transverse fields. The values are compared in Table II and the threshold field curve shown in fig. 2. From this curve the difference in specific heat of indium in the normal and superconducting states is calculated to be $S_n - S_s = 3{\cdot}48 \times 10^{-4}\,T$ cal./g.-atom (Daunt and Mendelssohn 1937). The Sommerfeld formula (Sommerfeld 1937) gives $S_n = 2{\cdot}97 \times 10^{-4}\,T$ cal./g.-atom.

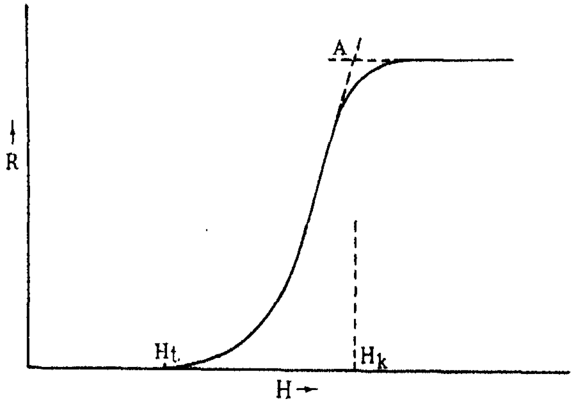


FIG. 1. Method of determining H_k from transverse field transition curve.

3. RESULTS

(a) *Tin*

The metal was supplied by Hilger (Lab. No. 10,000). Three specimens were investigated; Sn V, $R_{4\cdot2}/R_0 = 6{\cdot}6 \times 10^{-4}$; Sn VII, $R_{4\cdot2}/R_0 = 6{\cdot}7 \times 10^{-4}$; and Sn VIII, $R_{4\cdot2}/R_0 = 6{\cdot}0 \times 10^{-4}$. The results obtained are shown in Table III. In the sixth column, ΔT is the temperature interval (° K.) below the normal transition point. Two of the transition curves are shown in fig. 3. In this graph the ordinates are resistances, and the abscissae are field strengths, plotted as fractions of H_k. The difference between the H_t's of the two curves is apparent.

(b) *Indium*

The metal was supplied by Hilger (Lab. No. 10,874). Two specimens were used; In I, $R_{4\cdot2}/R_0 = 3{\cdot}3 \times 10^{-4}$ and In III, $R_{4\cdot2}/R_0 = 3{\cdot}2 \times 10^{-4}$. The results are given in Table IV.

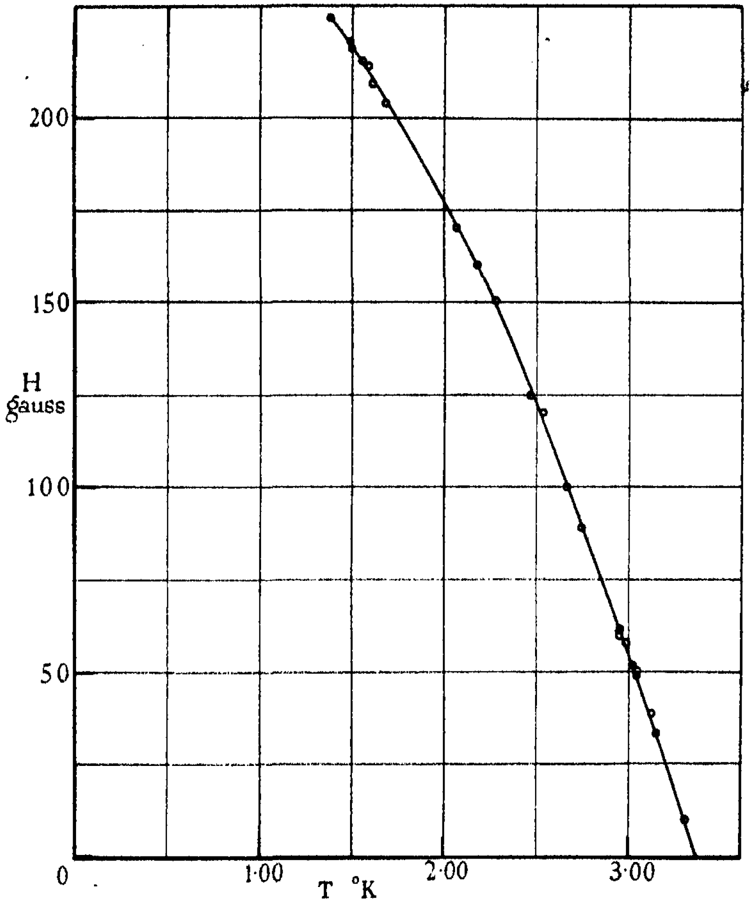


FIG. 2. Threshold field curve for indium.

TABLE III. TIN

Specimen no.	Temp. ° K.	H_t gauss	H_k gauss	H_t/H_k	ΔT ° K.
V	3·41	23	44	0·525	0·30
V	3·12	45	82	0·555	0·59
V	2·97	58	103	0·564	0·74
VII	2·92	63	110	0·572	0·81
VIII	2·80	72	124	0·580	0·91
VIII	2·75	76	130	0·585	0·96
V	2·71	78	135	0·591	1·00
VII	2·70	80	136	0·589	1·01
VIII	2·47	98	162	0·604	1·24
VII	2·29	115	182	0·634	1·42
V	1·49	174	248	0·704	2·22
VIII	1·49	175	248	0·706	2·22

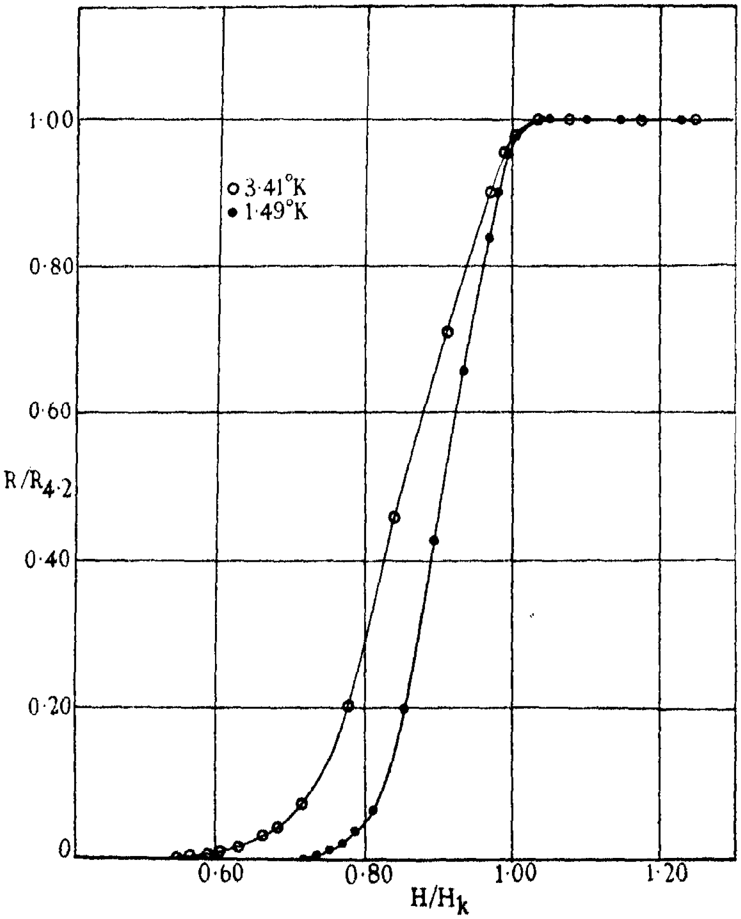


Fig. 3. Transition curves of tin (transverse field) at 3·41 and 1·49° K.

Table IV. Indium

Specimen no.	Temp. ° K.	H_t gauss	H_k gauss	H_t/H_k	ΔT ° K.
I	2·67	53	100	0·535	0·60
III	2·46	70	126	0·559	0·91
I	2·17	93	160	0·581	1·20
III	2·07	99	170	0·582	1·30
I	1·97	106	179	0·592	1·40
I	1·50	137	220	0·624	1·87
III	1·50	136	220	0·623	1·87

(*c*) *Lead*

The metal was supplied by Hilger (Lab. No. 8,334). Three specimens were measured; Pb III, $R_{4\cdot2}/R_0 = 9 \times 10^{-5}$; Pb IV, $R_{4\cdot2}/R_0 = 1\cdot8 \times 10^{-4}$; and Pb VII, $R_{4\cdot2}/R_0 = 4\cdot1 \times 10^{-4}$. Measurements were made at 4·2 and 1·68° K. only. The results are given in Table V.

TABLE V. LEAD

Specimen no.	Temp. ° K.	H_t gauss	H_k gauss	H_t/H_k	ΔT ° K.
IV	4·20	414	552	0·750	3·0
VII	4·20	412	552	0·746	3·0
III	4·20	416	552	0·754	3·0
IV	1·68	692	765	0·905	5·5
VII	1·68	697	765	0·913	5·5
III	1·68	696	765	0·910	5·5

The variation of H_t/H_k with temperature for tin and indium are shown graphically in fig. 4. From consideration of the probable sources of error it is estimated that the values of H_t/H_k are accurate to within 2 %, and it is seen that in the figure the straggling of the points from the best straight line

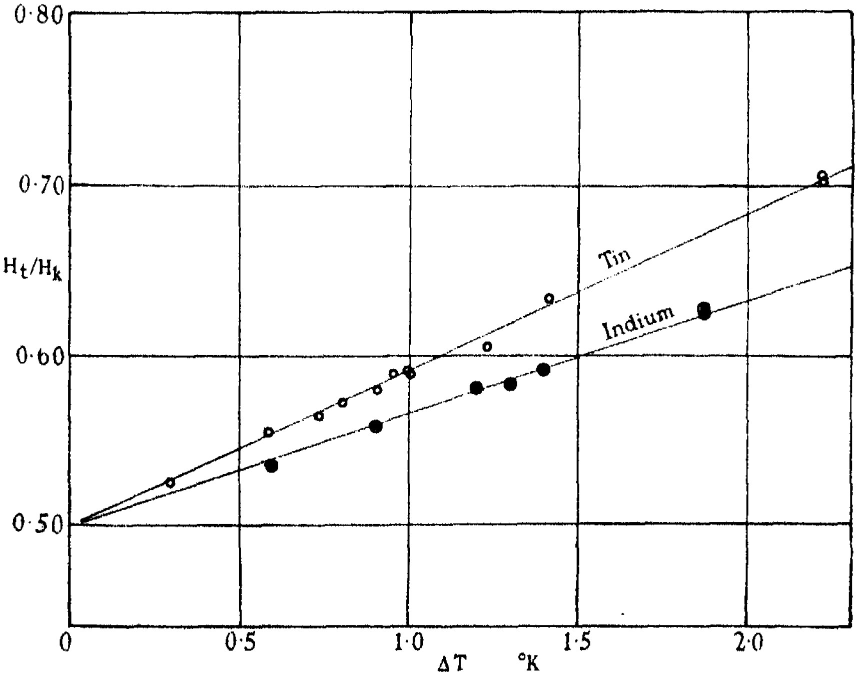


FIG. 4. Variation of H_t/H_k with temperature.

is within this limit. Both tin and indium give, by extrapolation from the experimental data, the value $H_t/H_k = 0{\cdot}500$ at the normal transition point ($H = 0$) of the superconducting metal.

It will be noted that the slope of the lines for tin and indium differ only slightly. The values for lead fall between these two lines. From this it is concluded that the variation of H_t/H_k with temperature depends very slightly, if at all, on the particular superconducting metal used. However, it is probable that the purity of the metal does exert an influence on this relation, and the differences in slope may be assigned to this cause.

4. DISCUSSION

In taking the observations the accurate setting of the potentiometer circuit required about 1 min. for each reading. After balance had been obtained no variation of the resistance with time was observed. Apparently the time effects found with the penetration of field into superconductors (Mendelssohn and Pontius 1936*a* and *b*; de Haas, Engelkes and Guinau 1937) do not occur in the restoration of resistance. Particular care was taken over this point when the applied field was between $0{\cdot}50\ H_k$ and H_t. In one instance the potentiometer remained accurately at zero ($R < 10^{-8}\,\Omega$) for 25 min., while the field applied to a tin specimen was $0{\cdot}58\ H_k$ and the temperature was kept at 2·70° K. Thus it is evident that the field which has penetrated into the cylinder has distributed itself in such a manner as not to destroy the perfect electrical conductivity of the specimen.

While these investigations were in progress, Mann and Wilhelm (1937) published the results of similar measurements made on a short single crystal tin cylinder using a persistent-current method. They found a similar linear variation of H_t/H_k with temperature. While their values of H_k for tin do not agree, at lower temperatures, with those found by other workers, the fact that both H_t and H_k are determined in the same manner for the same specimen makes it unlikely that their ratio will be much in error. Their results show that single crystals as well as pure polycrystalline cylinders show this phenomenon. The value $H_t/H_k = 0{\cdot}58$ given by de Haas, Voogd and Jonker (1934) from the transition curve of a single crystal wire at 2·92° K. ($\Delta T = 0{\cdot}81°$) also agrees well with the value obtained from fig. 4.

In attempting to explain why resistance does not appear when the field has penetrated the cylinder and why a larger proportion of the critical field is required to produce the first signs of resistance when the temperature is lower (i.e. when the critical field is increased), we may take either of two views as to what happens in the superconductor. If we imagine the field penetrating completely in layers of thickness of the order of 10^{-5} cm., as does Grayson-Smith (1937) in his discussion of Mann and Wilhelm's results, then it follows from London's theory that these "laminae" will have higher critical fields than the bulk metal. On the other hand, we may picture the field as penetrating through relatively large portions separated by still superconducting metal (cf. Mendelssohn and Pontius 1936*a*). In this picture we have no true "intermediate state" but simply a macroscopic mixture of normal and superconducting regions.

A preliminary experiment was performed to discover if it was possible to detect any evidence of the coexistence of these comparatively large normal and superconducting regions. A specimen 5 cm. long, cut from the same wire as Sn VIII (diameter 0·015 cm.), was laid on a thin bakelite plate and the usual current and potential leads soldered to it near the ends. At intervals on the central 12 mm. of this specimen, potential leads of 47-gauge eureka wire were bound to it and held in position by being passed through fine holes drilled in the plate. This type of mechanical contact was chosen since it produced the least distortion both in the cylindrical shape of the specimen and in the distribution of magnetic field in the immediate neighbourhood of the specimen. The potential leads were numbered 0, 1, 3, 4, 6, 12; the number of each contact corresponding to its distance, in millimetres, from the lead which was topmost when the specimen was mounted in the helium cryostat.

A transverse magnetic field (H) was applied which restored a portion of the resistance of the complete specimen ($H_t < H < H_k$). With temperature, external field and current flowing in the specimen all maintained constant, the resistance of the interval between each pair of leads was measured. The results are shown in fig. 5, the ordinates representing the percentage of normal resistance restored in the particular portion. The percentage of normal resistance restored by the applied field in the complete specimen (measured as a whole) is also shown. Graphs 1–4 were obtained at a temperature of 3·3° K. ($\Delta T = 0{\cdot}4°$), and graphs 5–7 at 1·7° K. ($\Delta T = 2{\cdot}0°$). The shorter intervals show a wide variation in the percentage of resistance restored, indicating that the size of the normal and superconducting regions is of the order of 1 mm. The value for the longer interval (6–12) agrees well, at $\Delta T = 0{\cdot}4°$, with the average of the specimen as a whole, showing that the size of the regions is small compared to the size of the interval (6 mm.). At the lower temperature, however (graphs 5–7), even this large interval may have a value quite different from the average of the complete specimen, indicating a larger size of normal and superconducting "grain" in this case.

From this direct evidence of the existence of measurable normal and superconducting regions in equilibrium we may form a picture of the behaviour of the cylinder when the external field is applied. We assume that when penetration commences at $H = 0\cdot50H_k$, the field enters the cylinder and forms normal regions in which the field strength is H_k. These

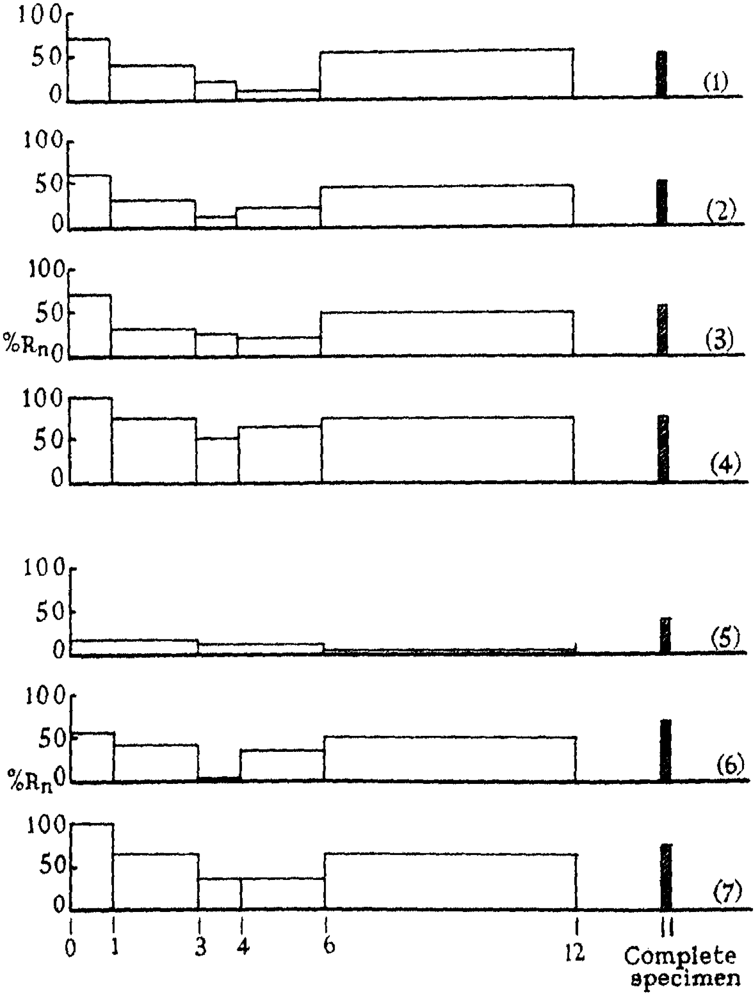


Fig. 5. Resistance of successive portions of a tin cylinder.

must be some sort of channels parallel to the direction of H, but their form will depend on the relation between the difference of energy of the normal and superconducting states ($H_k^2/8\pi$) and the energy of the surface of separation (Landau 1937). Consequently, their shape will probably be different when H_k has different values. When the critical field strength is small we suppose the normal regions to have a form something like that sketched in fig. 6*a*. When the critical field is large they may be more like fig. 6*b*. In both cases, as the applied field is increased, these normal regions will increase till they extend completely across the cylinder as shown by the dotted lines. Only when they have met to form a completely normal "layer" will the specimen begin to show electrical resistance. Because of the difference in shape, this will occur proportionately earlier in the case of low critical fields, and this accounts for the increase of H_t/H_k as H_k increases. Furthermore, the "layer" when formed at low temperatures (large H_k) is larger than when formed at higher temperatures, which is in accord with the results of the experiment described above.

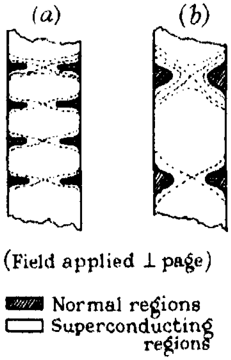


FIG. 6

This evidence for the coexistence of comparatively large normal and superconducting regions in a cylinder is in agreement with the opinion of J. G. Daunt (1937), based on results obtained with superconductors in superimposed D.C. and A.C. magnetic fields. Moreover, it suggests an explanation for the observed difference in his transition curves when the transverse D.C. field, used to destroy superconductivity, is applied (1) perpendicular to, and (2) parallel to the A.C. field which is used to indicate penetration (Daunt 1937, figs. 6 and 7). We have only to assume that the normal regions formed in the hollow tin cylinder by the D.C. field are of sufficient depth to penetrate the metal (1 mm. thickness). Then in case (1), D.C. field ⊥ A.C. field, since the normal regions form at the ends of a diameter perpendicular to the D.C. field, as soon as $H = 0{\cdot}50\, H_k$ completely "normal" paths will be presented to the A.C. field along the common axis of the primary and secondary coils. The number or size of these paths, and hence the flux in the secondary, will increase rapidly as the D.C. field is increased.

In case (2), D.C. field || A.C. field, the normal regions will be formed at the ends of a diameter perpendicular to the direction of the A.C. field. Consequently only a small amount of A.C. field will leak into the interior of the cylinder. The amount of flux will remain small even though the size or number of normal regions increases until there is a sufficient quantity of "normal" metal to provide a considerable channel in the direction of the A.C. field. Then the flux through the secondary coil will increase rapidly.

Since throughout a "normal" region the magnetic field strength is H_k, this simple picture also offers a possible explanation for the fact that in penetration experiments (1) the field as measured inside the cylinder with short bismuth wires sometimes exceeded the applied field in strength. A more detailed investigation, at present being undertaken, may give a better estimate of the size of these regions at different temperatures.

I wish to express my indebtedness to Dr J. D. Cockcroft for his advice and interest in this investigation and to Dr H. Jones for many discussions as to the probable interpretation of the results. Also to the Royal Commission for the Exhibition of 1851 for the award of a research scholarship during the tenure of which this work was carried out.

SUMMARY

Detailed experiments on the transition of polycrystalline cylindrical wires from the superconducting to the normal state in a uniform transverse magnetic field are described. Several specimens each of tin, indium and lead of high purity were investigated. The applied field strength (H_t) which restored the first traces of electrical resistance was accurately determined at various temperatures below the normal transition point. It is found that the ratio of this field to the critical field (H_k) corresponding to the temperature of the experiment is not constant but varies with the temperature. The variation is linear, the ratio H_t/H_k increasing as the temperature is decreased. For all specimens the ratio H_t/H_k would have the value 0·50 at the normal transition point.

It is considered that this effect is incompatible with the idea of an "intermediate state" existing in a long cylinder, and an explanation is offered based on the assumption that after penetration of the external magnetic field (above $0{\cdot}50\,H_k$) the cylinder breaks up into macroscopic regions of normal and superconducting material. The agreement of this hypothesis with previously observed phenomena is discussed.

A preliminary experiment is reported in which, by measuring the resistance of successive millimetre sections of a specimen, the coexistence of these normal and superconducting regions is shown directly.

REFERENCES

Daunt, J. G. 1937 *Phil. Mag.* **24**, 361.
Daunt, J. G. and Mendelssohn, K. 1937 *Proc. Roy. Soc.* A, **160**, 127.
Grayson-Smith, H. 1937 *Trans. Roy. Soc. Can.* (iii), **31**, 31.
de Haas, W. J. and Engelkes, A. D. 1937 *Commun. Phys. Lab. Univ. Leiden*, **4**, 325.
de Haas, W. J., Engelkes, A. D. and Guinau, O. A. 1937 *Commun. Phys. Lab. Univ. Leiden*, **4**, 595.
de Haas, W. J. and Casimir-Jonker, J. M. 1934 *Commun. Phys. Lab. Univ. Leiden*, **1**, 291.
de Haas, W. J., Voogd, J. and Jonker, J. M. 1934 *Commun. Phys. Lab. Univ. Leiden*, **1**, 281.
Kapitza, P. and Milner, C. J. 1937 *J. Sci. Instrum.* **14**, 165.
Landau, L. 1937 *Phys. Z. Sowjet.* **11**, 129.
von Laue, M. 1932 *Phys. Z.* **33**, 793.
London, F. 1936 *Physica*, **3**, 450.
Mann, K. C. and Wilhelm, J. O. 1937 *Trans. Roy. Soc. Can.* (iii), **31**, 19.
Mendelssohn, K. and Pontius, R. B. 1936 *a* *Nature, Lond.*, **138**, 29.
— — 1936 *b* *Physica*, **3**, 327.
Peierls, R. 1936 *Proc. Roy. Soc.* A, **155**, 613.
Schubnikow, L. W., Chotkewitsch, W. I., Schepelew, J. D. and Rjabinin, J. N. 1936 *Phys. Z. Sowjet.* **10**, 165.
Sommerfeld, A. 1937 *Ann. Phys., Lpz.*, **28**, 1.

Progressive lightning

V. A comparison of photographic and electrical studies of the discharge process

By B. F. J. Schonland,* D. B. Hodges† and H. Collens‡

(*Communicated by Sir Charles Boys, F.R.S.—Received* 5 *January* 1938)

[Plates 1-3]

The photographic method of studying the lightning discharge by means of the Boys camera has the unique advantage of giving direct information concerning events in the discharge in two dimensions of space and one of time and could be extended if necessary to include the third space dimension. The luminous events which it records are, however, secondary processes, and the primary movements of electrical charge which cause them can only be inferred by an application of ideas gained from the laboratory study of the passage of electricity through gases.

The direct study of these primary electrical processes involves the observation of the electric field during the discharge by means of a cathode-ray oscillograph. Studies of near lightning by this method have recently been reported by Appleton and Chapman (1937). The method gives information concerning the total electric moment of the cloud charges and requires to be compared with the photographic data before its results can be interpreted in terms of the charges themselves and their movements.

The present paper is concerned with field studies designed to correlate the information obtained by these two methods. It has proved possible to secure a certain number of direct and detailed correlations by simultaneous observations of the luminous and the electrical changes during a discharge. In general, however, the comparison of the two sets of results, electrical and photographic, is made from data drawn from different discharges. All the information is derived from storms in the neighbourhood of Johannesburg.

---

* The Bernard Price Institute of Geophysics, Johannesburg.
† Howard College, Durban.
‡ The Victoria Falls and Transvaal Power Co., Johannesburg.

1. Apparatus and method

The cameras employed in this work have already been described (Schonland, Malan and Collens 1935; Collens 1937). The instruments used for the electric field determinations were similar to those recently described by Watson-Watt, Herd and Lutkin (1937). An aerial was provided with a critical damping resistance and a capacity potentiometer with a known shunt resistance to ground. An amplifier with resistance-capacity coupling obtained its input from the aerial condenser (fig. 1). With suitable values for the capacities and resistances employed the output of this arrangement was directly proportional to the electric field at the aerial. The output was transferred to a cathode-ray oscillograph and recorded on a camera with a moving film.

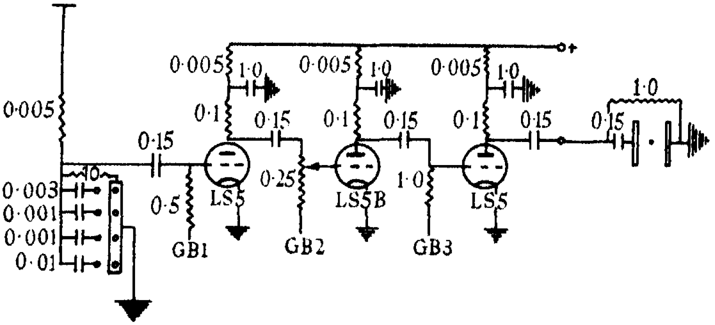


FIG. 1. Capacities in microfarads, resistances in megohms.

The oscillograph was a Cossor gas-focused tube with type J screen and supplied with 2000 V. from a battery of dry cells. Records were made on Kodak sound-film through a Dallmeyer $f/1\cdot5$ lens of focal length 5 cm.

The recording film was 102 cm. long and was carried on the circumference of a rotating drum provided with the axial traversing motion described by Lutkin (1937). The drum was rotated by an electric motor at about 80 rev./min., the speed being checked during each run. This method causes the zero position of the recording spot to describe a spiral on the film, and a mask over the zero point on the face of the oscillograph prevents this zero spiral from being recorded. The pitch of the spiral was made 0·6 mm., so that the film had an effective length of 170 m. and ran for about 2 min. The record of a discharge with a total duration of 0·5 sec. thus occupied some 70 cm. of film length and could be read easily to 0·1 mm. or 70μsec.

To make satisfactory use of this direct and economical system of recording it was necessary to keep the recording spot on the masked zero line during the periods when actual flashes were not occurring. For this reason no attempt was made to study the slower field changes which occur in the intervals between separate flashes nor those which occur between separate strokes of the same flash. The aerial amplifier system was arranged to have an overall time constant ranging from 0·03 to 0·01 sec. according to the capacity of the aerial condenser (fig. 1) which was employed. Slow and semi-permanent field changes occupying times of this order were therefore distorted, the recording spot returning or tending to return to zero. For shorter times, up to the limit of resolution of the arrangement, 50μsec., the amplifier was found by tests with artificial transients to be giving an undistorted response. The total amplification used was 260 times.

The aerial usually employed was a wire 2·4 m. long carried on sulphur insulators protected against rain and held at 33 cm. above the roof of the lorry. An additional aerial was required for storms at distances exceeding 20 km.

The instruments were mounted inside a motor-lorry which could be driven to a point in the path of an approaching thunderstorm. All the oscillographic records were taken in conjunction with either visual or photographic observations of the discharges which caused them, a system of bell signals serving for communication between the outside observer and the oscillograph operator, who had means of marking the film in a suitable manner. Thus each field change record has been definitely correlated with the flash which caused it, and in a number of cases a Boys camera photograph of the discharge is also available.

2. RESULTS OF COMPARISON OF ELECTRICAL AND PHOTOGRAPHIC RECORDS

Records of the field changes produced by seventy-one discharges to ground and a much larger number of discharges within the cloud have been obtained with this equipment. The data were gathered from more than thirty different thunderstorms.

The greater portion of this paper is concerned with discharges to ground, which it is possible from both oscillographic and photographic information to divide into two main classes.

The first class, which we term type α, gives an oscillograph record of its electric field effects which is illustrated diagrammatically, but not to scale, in fig. 2A.*

The illustration shows that three major changes of field occurred during the discharge, each of which is separable, as was first shown by Appleton and Chapman, into three parts, *a*, *b*, and *c* as marked in fig. 2. These authors have given reasons for identifying the *a* portion with the leader process, the *b* portion with the rapid return streamer, and the *c* portion with the continuance of luminosity in the channel after the return streamer has reached the cloud. It is the object of this section to confirm and extend these identifications.

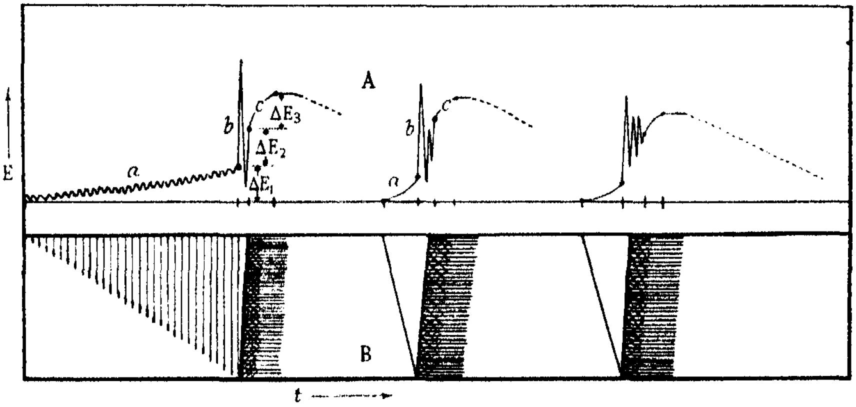


Fig. 2.

The records of three actual discharges to ground are shown in fig. 3 (Plate 1), each consisting of three major field changes.

Fig. 2B illustrates the type of picture which the identifications to be established later indicate as given from the same discharge on a camera with a fixed lens and a film moving at the same linear velocity as that used for fig. 2A. The picture is similar to one previously discussed (Schonland, Malan and Collens 1935), and is drawn for the case of a vertical discharge channel, so that displacements in the horizontal direction are displacements in time. About 65% of all discharges to ground give photographic and field-change records of this type.

---

* The slow field variations between the three major field changes are not shown on this diagram. The dotted lines indicate how the time constant of the amplifier affects the record in the intervals between strokes. Recent studies show that the changes during these intervals are unimportant.

A variation in the nature of the first, stepped, leader process which is illustrated in fig. 4B, is found to give rise to a significant variation in the corresponding record of the first major field change, as shown in fig. 4A. The stepped process for this type β leader is rapid in its earlier stages and advances with long bright steps (Schonland 1938), but later on its effective velocity, its step length and brightness become very much smaller, and it can only be photographed on a Boys camera with much difficulty. It will be shown that the effect of the later portion of this leader upon the electric field is a small and slow change of field strength. For this reason the comparatively small time constant of the amplifier causes the recording spot of the oscillograph to fall to zero after the point P of fig. 4A. About 35% of the flashes to ground which we have studied have leaders of this nature.

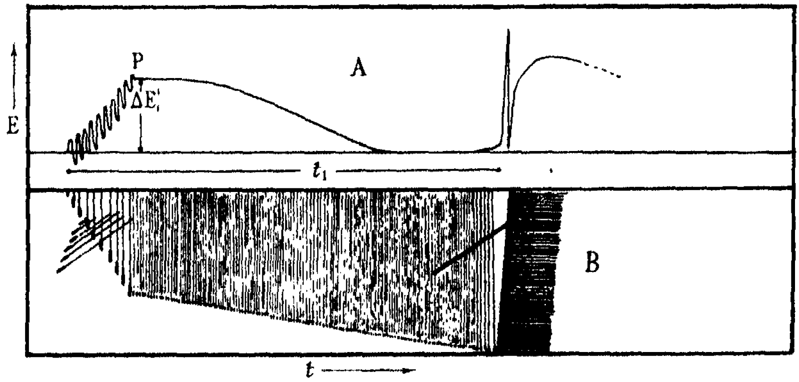


Fig. 4.

A number of examples of records of the first strokes of type α discharges are shown in fig. 5 (Plate 1). Fig. 6 (Plate 2) begins with two records of type β,[*] the remainder of the figure being occupied by the first and strongly rippled portions of various others of the same type.

---

[*] The crossing of the zero line during the later portion of the a field change in the first record of fig. 6 is a common feature of such records and due to the properties of the amplifier used. An amplifier of longer time constant does not show it.

(*a*) *Identification of major field changes with successive strokes of the discharge*

This identification is most clearly established from simultaneous observations by the photographic and the electrical methods. The correspondence in number and in time separation of major field changes and photographically observed strokes has been established in the case of four lightning flashes, containing six, three and four strokes respectively. An example is shown in Table I below. The Boys camera and the oscillograph drum speeds were independently determined, and the agreement is well within the limits of error for such determinations.

TABLE I. FLASH Y 3

Stroke	Time of start of return streamer on camera sec.	Time of start of *b* field-change on oscillograph sec.
1	0	0
2	0·155	0·150
3	0·084	0·082
4	0·046	0·044

In view of this correspondence and as further support for it, it is of interest to compare the distribution curve for the time intervals between major field changes due to flashes known to pass to ground with that for the time intervals between photographically observed strokes, obtained independently. This comparison is shown in fig. 7 which indicates that in both cases the most frequent interval lies between 0·03 and 0·04 seconds.

(*b*) *Identification of the* a *field change with the leader process*

The first leader to all flashes examined photographically under good conditions has been found to be stepped (Schonland, Malan and Collens 1935). It is therefore to be expected that the *a* portions of all first-stroke field changes will show evidence of electrical pulsations due to the sudden appearance of each step streamer. These pulsations have been reported by Appleton and Chapman (1937) though without reference to the order of the major field change considered. They are observed by us on practically all the *a* portions of the first major field changes. An examination of twenty-one of these with sufficient resolution to permit of measurement of the average time interval between pulsations shows that this quantity ranges from 40 to 65μsec. This is in satisfactory agreement with the photographically observed interval between steps of the first leader process which has been shown (Schonland, Malan and Collens 1935) to range from 31 to 91μsec. Some observations of the same kind have been reported by Appleton and Chapman (1937).

It is significant that pulsations of this kind are not usually shown on the *a* field changes of subsequent strokes, as figs. 2, 3 and 8 indicate. This is in accord with the absence of steps, except occasional "dart-steps" (Schonland, Malan and Collens 1935) too rapid to be recorded by our equipment, on the photographic records of the dart leaders to subsequent strokes.

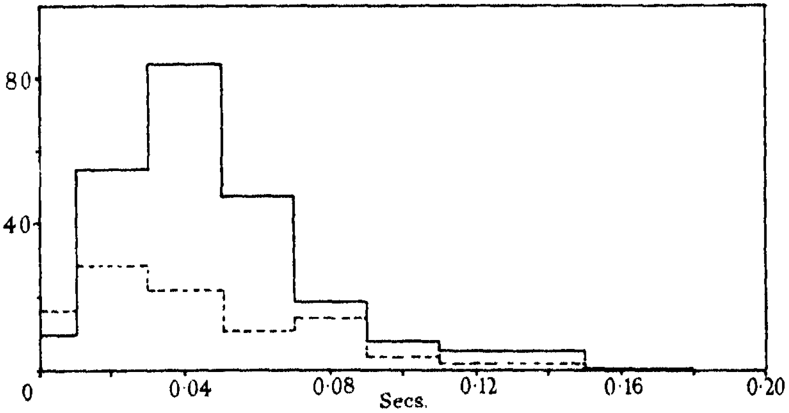


FIG. 7. Distribution of time intervals between major field changes (continuous line) and between photographically observed strokes (broken line).

Definite identification of the *a* field change with the leader process has been obtained from simultaneous camera and oscillograph records of the two subsequent strokes shown in fig. 8*a*, *b* (Plate 3). Stroke Y 3_2 gave a leader which is visible on the record from the fast lens (top left)* for 170μsec. (fig. 8*a*). The corresponding oscillograph record shows an *a* field change lasting 240μsec. Stroke Y 3_3 gave a leader which was too slow to be reproduced from the fast lens record but which is shown on the slow lens record as a faint line to the left of the return streamer (fig. 8*b*). In this case the duration of the luminous leader process was 1900μsec., while the corresponding *a* field change lasted 3000μsec. The longer durations found for the field changes in these cases are no doubt due to the ability of the oscillograph to record earlier portions of the leader process which are hidden from the camera by the cloud.

---

* Inadvertently reversed in printing.

Both Y 3_2 and Y 3_3 were subsequent strokes; the leader to the first stroke of the series, Y 3_1, was too slow and feeble to be recorded on the Boys camera.

In the case of such first strokes we have not as yet obtained a simultaneous direct correlation by the two methods. One indirect means of establishing the identification, by comparing the pulsations on the *a* portion with the leader steps, has already been discussed. Another is provided by a comparison of the durations of leaders and of *a* field changes. This comparison is made for first strokes in fig. 9, where the curve *A* gives the distribution in duration of sixty-nine *a* field changes from different discharges, forty-three of type α and twenty-six of type β. In the latter case the duration is measured as t_1 in fig. 2 A, and the contribution made by these type β discharges is shown as a shaded area in the distribution curve. Fig. 9 B shows the corresponding curve for the duration of thirty-two stepped leaders observed photographically. Here the shaded parts indicate the contribution made by eleven leaders of type β which showed the changes in effective velocity and brightness already described.

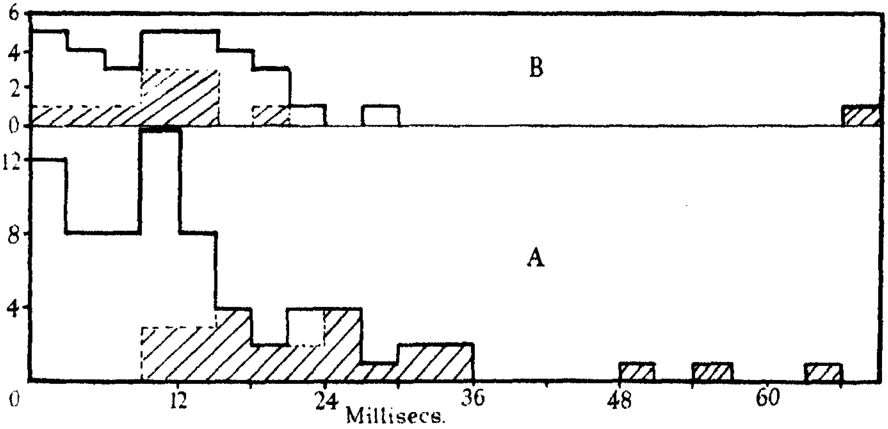


FIG. 9. A. Duration of *a* portion of first stroke field-changes. B. Duration of first leader processes determined photographically. Type β leaders shaded.

For strokes subsequent to the first the times occupied by the leader and *a* processes are very much shorter, as shown by the distribution curve of fig. 10 which refers to 114 *a* changes in subsequent strokes. The peak of the curve is at 1 msec. in agreement with what is to be expected from the photographic data, for the most frequent value of the velocity of a dart leader is $2 \cdot 7 \times 10^8$ cm./sec. and thus a path length of 3 km. would involve a duration of the order found.

Before leaving the leader portion of the field change we may note a further point connected with the pulsations superimposed upon it. The duration of this pulsating portion in the case of a type α leader is the whole duration of the leader (fig. 2), but in the case of a type β leader it is difficult to observe the pulsations after the first rapid stage of the leader process is over (fig. 4). In order to establish the point that this strongly pulsating portion is associated with the first and vigorous stage of the photographically observed type β leader the durations in the two cases must be compared. This is done in fig. 11, the *A* curve of which shows the distribution in duration of the strongly pulsating portion of seventeen type β leader field changes, while the *B* curve gives the distribution in duration of the vigorous part of ten such leaders observed photographically. The agreement in order of magnitude is satisfactory, about 1 msec., and the fact that the peak of the *A* curve lies at a higher value than that of the *B* curve is further evidence for the view that a considerable part of the leader process is hidden within the cloud.

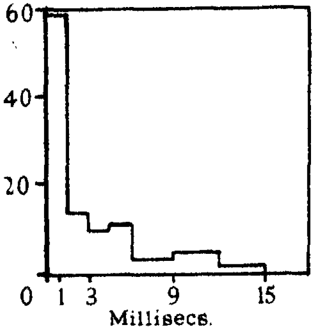


Fig. 10. Duration of *a* portion of subsequent stroke field changes.

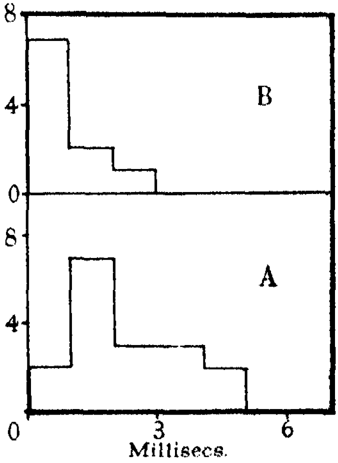


Fig. 11. A. Duration of strong superimposed pulsations on type β leader field changes. B. Duration of vigorous stepped portion of leaders observed photographically.

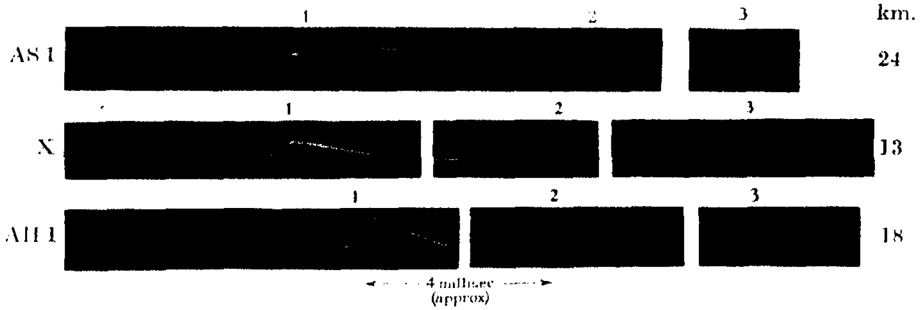

1
2
3
km.
AS 1
24
1
2
3
X
13
1
2
3
AH 1
18
4 millisec
(approx)

Fig. 3

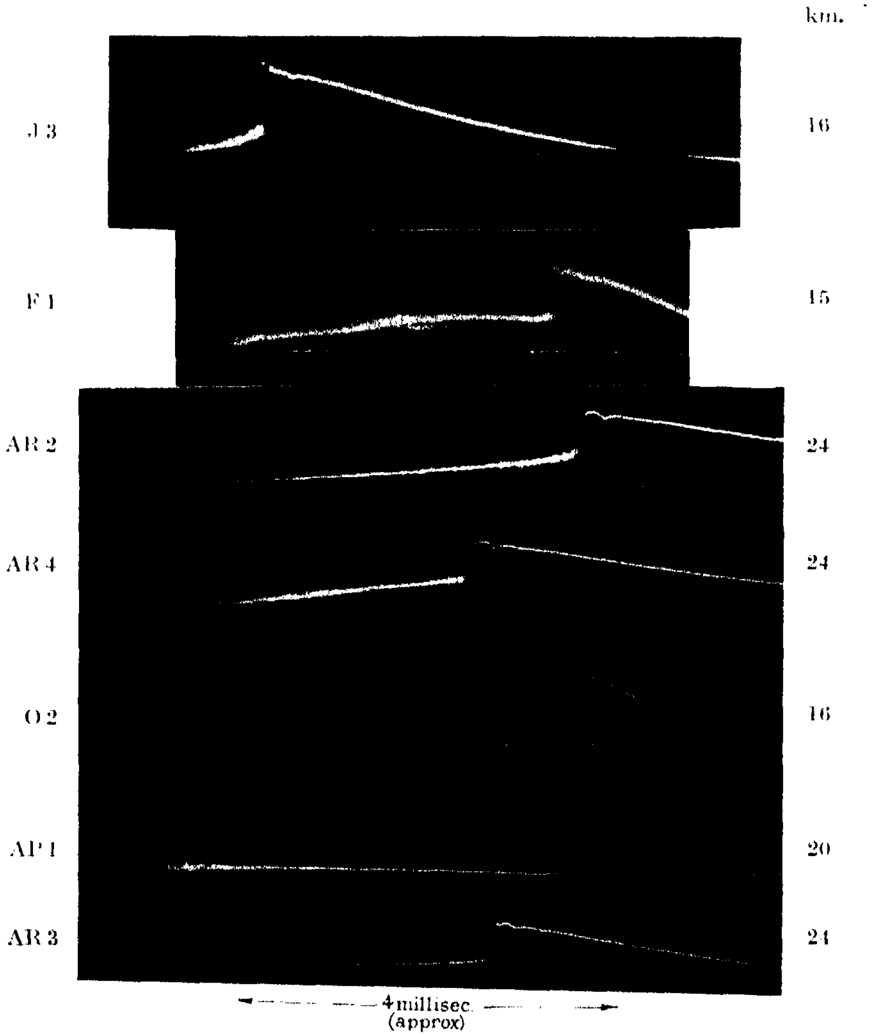

km.
J 3
16
F 1
15
AR 2
24
AR 4
24
O 2
16
AP 1
20
AR 3
24
4 millisec.
(approx)

Fig. 5

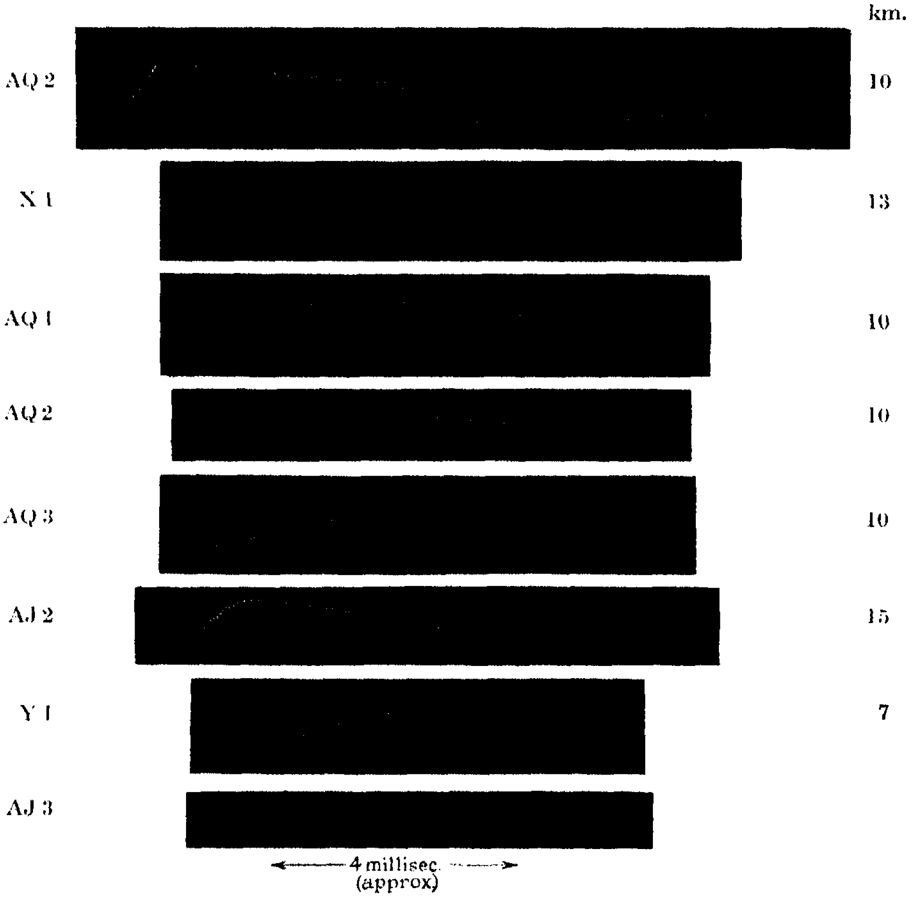

km.
AQ 2
10
X 1
13
AQ 1
10
AQ 2
10
AQ 3
10
AJ 2
15
Y 1
7
AJ 3
4 millisec.
(approx)

FIG. 6

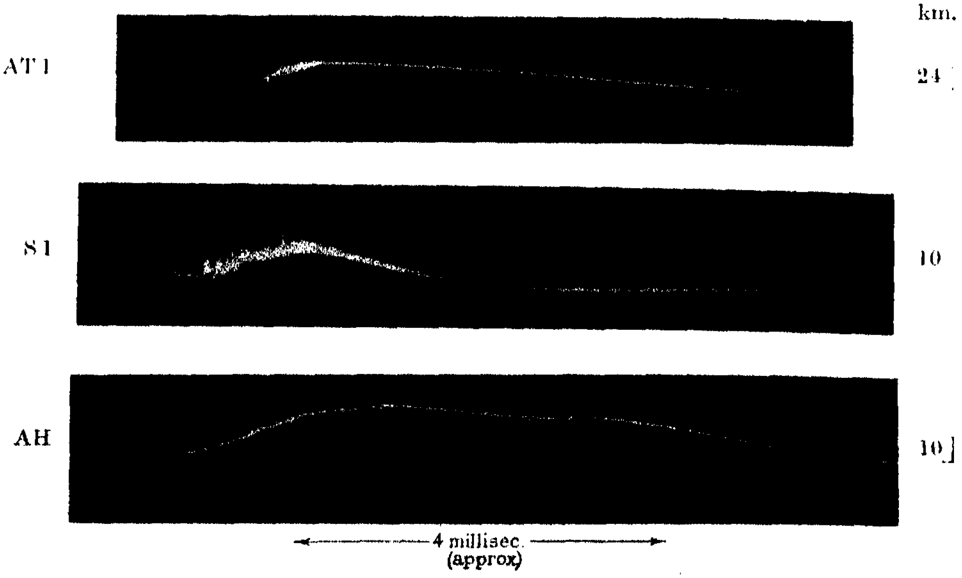

km.
AT 1
24
S 1
10
AH
10
4 millisec.
(approx)

FIG. 17

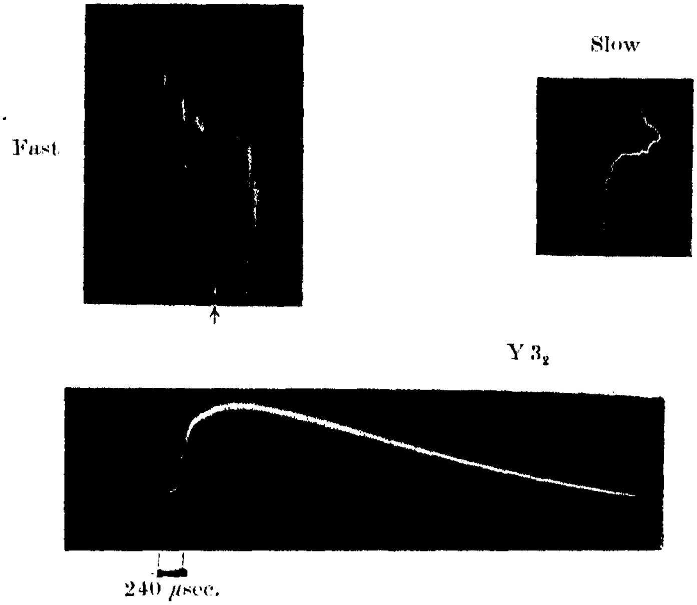


Fig. 8*a*. Fast lens. Leader time 170 μsec.

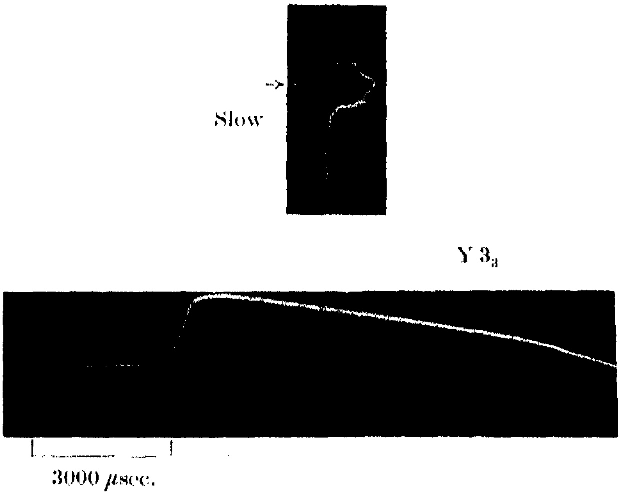


Fig. 8*b*. Slow lens. Leader time 1900 μsec.

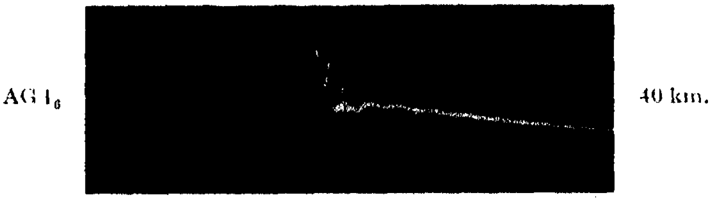


Fig. 14

(*c*) *The* b *and* c *field changes in the case of first strokes*

The *b* part of the oscillograph record follows immediately upon the leader or *a* part and involves a nett field change of the same order but produced in a much shorter time. Superimposed upon this is a pulsation of some complexity, whose general nature is shown in fig. 2.

This *b* change can be identified with the effect of the rapid passage of the return streamer up the channel formed by the leader, discovered by Schonland and Collens (1934). This identification is supported by the simultaneous camera and oscillograph observations shown in Table II below.

TABLE II

Stroke	Duration of *b* change on oscillograph μsec.	Duration of return streamer process on camera μsec.
Y 3_4	75	51 ± 10
Y 3_1	150	110 ± 30
Y 2_1	140	60 ± 40

Since it is difficult to determine the full duration of the rapid return streamer process, which involves passage of the streamer along the branches as well as into the cloud itself, the general agreement shown in the table may be considered satisfactory.

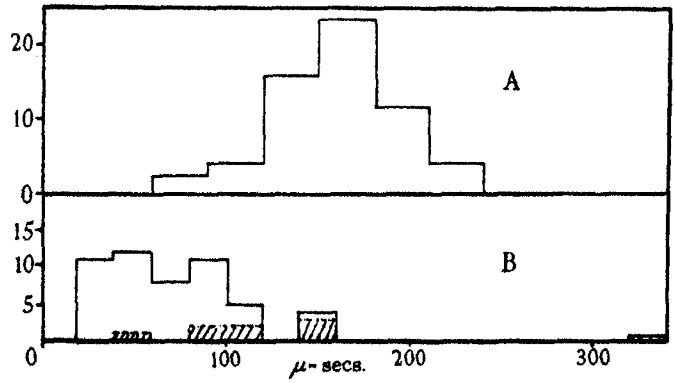


FIG. 12. Durations of *b* changes (A) and of return process (B).

Further support for this identification is given by a comparison of the durations of the *b* field changes of sixty first strokes shown in fig. 12 A with the duration of the rapid return streamer process in the case of fifty-six photographic observations of first strokes with the Boys camera. This latter duration should include the time taken for the streamer to reach the end of its last and longest branch, and where this has been determined the duration is shown as a shaded area in the figure. The first curve has its peak at 165μsec., the second ranges from 20 to 160μsec. The shaded areas of the B curve suggest that if these photographs could all have been analysed to give the duration of the complete discharge of the branched leader system by the return streamer, the agreement would have been much closer. The majority of the times in this lower figure refer to the time taken by the streamer to reach the cloud base only, and it is therefore to be expected that they would be lower than the true duration of the return process.

It is natural to suppose that the final c field change corresponds, in part at least, to the discharge to ground of that portion of the cloud charge which was not lowered into the air by the leader, and which in its slow passage to earth keeps the channel luminous after the return streamer stage is over. The duration of the c change in the case of thirty-nine first strokes has been found to range from 70 to 900μsec. Malan and Collens (1937) state that the most frequent value for the duration of luminosity in the channel is found photographically to be 1000μsec. Since their observations show that this quantity ranges from a few hundred μsec. to a full half-second, no detailed comparison is possible. Fig. 13, which gives the distribution in duration of the c change from these thirty-nine first strokes, indicates that the greater portion of the stroke is completed with a c change of less than 300μsec. duration.

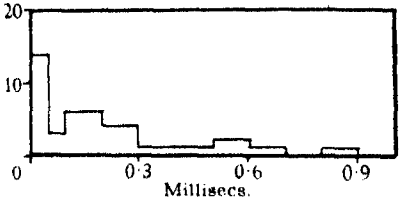


FIG. 13. Duration of final c field change.

(*d*) *The* b *and* c *changes for subsequent strokes*

In a subsequent stroke field-change the change which follows the a portion is unexpectedly complicated and cannot always be separated into a b and a c portion, for the pulsations superimposed upon the electrostatic field change are much more complex and of longer duration than that found in the case of a first stroke. A general picture of what is observed is shown in fig. 2 A, and examples are given in fig. 3 (Plate 1) for the second and third strokes of three discharges and in fig. 14 (Plate 3) for the sixth stroke of another discharge. The latter is typical of the complexity which can develop and which is not observed on first strokes. It is found that in the majority of the flashes examined this complexity increases progressively from stroke to stroke of the series which make up a flash. It is rare to find a subsequent stroke with superimposed *b* pulsations of less complexity and shorter total duration than those on the first stroke.

The unexpectedness of this effect lies in the fact that the camera discloses no such complexity in either the path of a subsequent stroke return streamer or the after-glow which succeeds the return streamer. It must therefore be ascribed to discharge processes within the cloud, which cannot be observed by the camera and which take place after the return streamer has entered the base of the cloud. Such processes must be associated with the suggestion (Schonland 1937) that leaders to subsequent strokes originate in discrete centres of charge in different parts of the cloud. The increased duration of the pulsating *b* field change would on this view be due to the increased distance the return streamer penetrates into the cloud, an effect similar to that already noted for the leader streamers studied in § 2 (*b*). The increased complexity would be ascribed to the development of new and rapid discharge processes within the cloud itself and the progressive nature of the effect to the fact that as a rule each successive stroke taps a centre of charge at a greater distance from the original channel.

(*e*) *The polarity of the discharge to ground*

The oscillographic material at our disposal includes records of the field changes due to seventy-one discharges, from some thirty different thunderclouds, which are definitely known to have passed from cloud to ground. It also contains records of several hundreds of discharges which are known to have occurred within the cloud. The ground flashes comprise 350 separate strokes, each of which causes a major field change which is divisible, as has been seen, into an effect due to the downward leader and another due to the return process. The nett electrostatic change for each stage of every flash to ground thus recorded has been found to be positive, indicating without ambiguity that each and every stage has involved the slow lowering and sudden destruction of a negative charge from the cloud.*

No such regularity is observed in the sign of the field changes from discharges which take place within the thundercloud. The separate strokes of a cloud discharge show a reversal of sign of field change with distance of exactly the same nature as has previously been noted for the nett change due to the whole discharge.

---

* A single exception has recently been observed in the course of further work.

3. THE NATURE OF THE ELECTROSTATIC FIELD CHANGES IN THE DISCHARGE TO GROUND

(a) *The form of the leader field-change record*

An approximate description of the form of the electrical record produced by the leader process can be obtained theoretically from the assumption that the stem of the leader streamer carries a uniform charge, q, per unit length and moves with uniform velocity, v. In the case of the first leader this velocity may be taken to be the effective velocity of advance, and the step streamers may be left out of account. The leader process then lowers charge downwards from the cloud in order to maintain newly formed sections of the stem in a charged condition.

Such a leader streamer as has been postulated above, which proceeds vertically downwards from a cloud centre at a height H above the ground, will have reached a height $H-vt$ after time t and will carry a charge qvt upon its stem. The total electric moment of this charge and its image is $2qvt(H-vt/2)$, and this will produce a vertical field at a point on the ground at a distance L, supposed large compared with H, which is of strength $2qvt(H-vt/2)/L^3$. Since the charge qvt was initially at a height H the resulting change of field is given by

$$\Delta E = 2qvt[H-(H-vt/2)]/L^3 = qv^2t^2/L^3. \qquad (1)$$

If, as we have supposed, both q and v are constant, $\Delta E \propto t^2$, and the form taken by the field-change record of the leader process will be a parabola which is concave upwards. This parabolic curve is a characteristic feature of the a field change and is shown in a number of cases reproduced in (figs. 5 and 8a, b, Plates 1, 3).

For a straight leader channel at an angle to the vertical the same treatment yields $\Delta E = qv^2t^2 \cos\theta/L^3$, which is again a parabolic relation. Thus, as is otherwise obvious, a leader which begins with a channel which is practically horizontal will produce an initial field change which is small compared to that due to its later and more nearly vertical channel. Discharges with a flat top to their channel are quite often observed and a field changes of corresponding nature are frequently found.

The characteristic feature of what we have termed a type β leader is an abrupt alteration in the value of its velocity after it has reached a certain space-charge region in the air below the cloud. Its initial high velocity may reasonably be associated with a high value of the stem charge q, and hence eqn. (1) leads one to expect an initial leader change which is parabolic and of large curvature, as is shown by the examples in fig. 6 (Plate 2). The subsequent downward movement is observed photographically to be much slower and this is associated with a much smaller value of q. Hence the curvature of the field-change record after this point may be expected to be small, as in the second example in fig. 6, and often so small that the time constant of the amplifier causes the return of the recording spot to zero, as in the other examples in this figure.

The effect produced by tortuosity in the channel may be examined in terms of the case where the leader travels vertically for a time t_1 and then moves at an angle θ to the vertical for a further time t_2. If q and v are again supposed constant the field change after a time t_1+t_2 is given by

$$\Delta E = qv^2(t_1^2 + 2t_1t_2 + t_2^2 \cos\theta)/L^3.$$

In practice θ does not often exceed $\pi/2$, and in this extreme case

$$\Delta E = (qv^2t_1^2 + 2qv^2t_1t_2)/L^3 = [\Delta E]_{t_1} + \{d/dt[\Delta E]_{t_1}\}t_2 \qquad (2)$$

Thus if the streamer turns abruptly from the vertical to the horizontal at t_1, the further field change after this will follow a straight line tangential to the parabola at t_1. For a turn of less than $\pi/2$ the further field change will be intermediate between this parabola and its tangent at t_1.

Such abrupt alterations in direction occur in the leader process as a result of the steps in the first leader (Schonland, Malan and Collens 1935) but their electrostatic effects, together with the superimposed pulsations due to induction and radiation field changes, require instruments of higher resolving power than we have used to be properly studied. In the return streamer they occur at intervals of the order of 7 μsec. and have not yet been observed.

(*b*) *The distribution of charge along the leader channel*

In the foregoing we have assumed a uniform distribution of charge along the leader channel. A test of this assumption can be made by comparing the final leader field change ΔE_1 (fig. 2A) with the nett electrostatic change in the *b* or return streamer stage, ΔE_2.

Let Q be the total charge distributed along the leader channel at the moment it reaches the ground, and H/n its effective height above the ground.

For a straight channel and uniform charge per unit length, $n=2$. It is easily shown that the total leader field change is given by

$$\Delta E_1 = 2QH[(H^2+L^2)^{-\frac{3}{2}} - (H^2/n^2+L^2)^{-\frac{3}{2}}/n], \tag{3}$$

which when $L \gg H$ becomes

$$\Delta E_1 = (2QH/L^3)\,(\overline{n-1}/n).$$

The leader field change, which effectively lowers Q from H to H/n, is followed by the return or b field change which lowers it from H/n to ground and is given by

$$\Delta E_2 = 2QH/nL^3 \quad \text{for } L \gg H.$$

From these two equations we have

$$\Delta E_2/\Delta E_1 = 1/\overline{n-1}, \tag{4}$$

an equation from which n can be determined experimentally.

In applying eqn. (4) to our present data care has to be taken that the effect of amplifier distortion does not enter into the measurement of the slow leader field change ΔE_1, and only those field changes can be used which take place in a time too short for the distortion to be appreciable. The more rapid type α leaders and the leaders to subsequent strokes are suitable for this purpose.

The distribution of the values of $\Delta E_2/\Delta E_1$ in twenty-six type α first strokes and forty-six subsequent strokes is shown in fig. 15. The subsequent strokes selected were those in which the b portion of the field change was not particularly complex and could be distinguished from the c portion without difficulty. It will be seen that the most frequent value for the ratio is 1·0. This indicates that the most frequent value of n in eqn. (4) is 2·0, and that the effective height of the charge distributed along the completed leader channel is usually $H/2$. We have seen that this is what would be expected for uniform leader charge distribution along the channel.

Of the values shown in the figure, 85% lie between 0·3 and 3·0, corresponding to effective heights for the lowered leader charge ranging from $H/4$ to $3H/4$. The more extreme values are found with first-stroke leaders where the effect of branching may be expected to produce considerable variation in the ratio.

It appears from these results that it is permissible, to the rough degree of approximation required in the treatment of the lightning discharge, to regard the leader process in general as involving a uniformly charged channel.

(c) *Type β leaders*

In the type β leader, the first part of the field change, $\Delta E_1'$ (fig. 4 A) is generally abnormally large, which indicates that the leader at this stage, as would be expected from its velocity and the vigour of the steps it executes, is lowering an unusually large charge. The return field change, ΔE_2, is found to be usually about 1/10th of $\Delta E_1'$, and in no case has it been observed to exceed 0·7 $\Delta E_1'$. It would follow that the charge lowered by the first part of the leader process is either placed close to the ground in the first stage or is no longer present when the return streamer starts back. The photographic evidence lends no support to the first explanation, for all type β leaders reach the end of their vigorous stage less than halfway down to the ground. The second explanation is in accord with the view (Schonland 1938) that the charge carried down in the first stage is neutralized by a space charge in the air, to which it was originally directed.

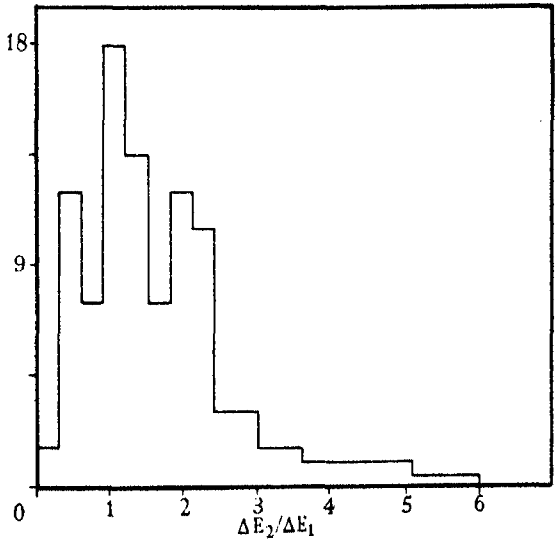


FIG. 15. Ratio return field change/leader field change.

(d) *Reversal of field change in the leader process*

Equation (3) above, which involves no particular assumption as to the distribution of charge along the leader channel, indicates that at a certain distance L_0 from the discharge the sign of the final leader field change ΔE_1 will reverse after passing through zero values. The value of L_0 depends upon H and n. For $n = 4/3$, 2·0 and 4·0, L_0 is 1·27 H, 1·01 H, and 0·75 H. In our storms this reversal distance should be about 3 km. It has been observed that within distances of this order the expected reversal occurs, but all the material discussed in the present paper was obtained at distances exceeding $2L_0$ for which the approximations used are justified.

(e) *The charge remaining on the cloud after the leader process has reached the ground*

The field change due to the combined action of the leader and the rapid part of the return stroke lowers a charge Q to the ground and is given by

$$\Delta E_1 + \Delta E_2 = 2QH/L^3, \quad \text{if } L \gg H.$$

If the whole of the final c field change is due to the subsequent passage to ground of a charge Q', which was left by the leader in the original centre of charge in the cloud, this final change is given by

$$\Delta E_3 = 2Q'H/L^3 \quad \text{for } L \gg H.$$

Hence

$$Q'/Q = \Delta E_3/\overline{\Delta E_1 + \Delta E_2},$$

or

$$Q'/\overline{Q+Q'} = \Delta E_3/\overline{\Delta E_1 + \Delta E_2 + \Delta E_3},$$

an equation from which the fraction of the original charge left in the cloud after the leader process has reached the ground can be estimated.

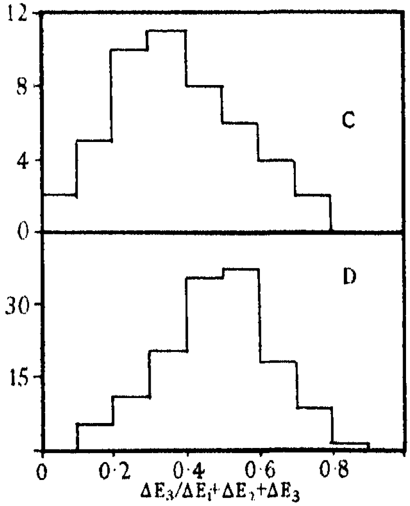


FIG. 16. Ratio $\Delta E_3/\overline{\Delta E_1 + \Delta E_2 + \Delta E_3}$; C, first strokes; D, subsequent strokes.

A distribution curve for the values of this fraction in the case of forty-eight first strokes is shown in fig. 16C. The most frequent value is 0·3, indicating that at most 30% of the cloud charge remains behind.

A similar curve for 152 subsequent strokes is shown in fig. 16D. In this case the most frequent value for the ratio is 0·5, indicating that 50% of the cloud charge remains behind. The leaders in this case are dart leaders and unbranched, and the increased value of the fraction left behind indicates that they are less effective in lowering charge than the first and heavily branched leaders.

4. TRANSIENT RADIATION AND INDUCTION EFFECTS ON THE OSCILLOGRAPH RECORD

We have so far dealt only with the electrostatic field, E_s, and its alterations during the discharge process. The full expression for the field recorded on the oscillograph includes the two transient induction and radiation terms, E_i and E_r, and is given by

$$E = E_s + E_i + E_r = M \,.\, 1/L^3 + dM/dt \,.\, 1/cL^2 + d^2M/dt^2 \,.\, 1/c^2L,$$

where M is the total moment of all the cloud charges at time $t - L/c$.

From these equations

$$E_i/E_s = \left[\frac{1}{M} \,.\, dM/dt\right] \times L/c \quad \text{and} \quad E_r/E_s = \left[\frac{1}{M} \,.\, d^2M/dt^2\right] \times L^2/c^2.$$

Since in the observations discussed in this paper the quantity L/c is of order 5×10^{-5}, the transient terms are not likely to be very large unless the quantities in square brackets are large, that is, unless the discharge process involves large and sudden changes of moment. The photographic evidence indicates that such changes must be expected during the execution of the rapid steps in the first leader and during the rapid return streamer process. It has already been noted that these are the two portions of the discharge which show powerful transient and pulsating effects on the electrical record. It is not proposed to discuss them in detail in the present paper, since the resolution of the oscillograph used was not sufficient to obtain full information on these transient effects, but a few general points may be noted.

The pulsations on the record of the stepped leader process show a time separation which corresponds well with the intervals of time between the starting of successive step streamers and a form which is easily interpreted as a combination of the induction and radiation terms resulting from the rapid lowering of charge in a step. The exceptionally large amplitude of the pulsations carried on the first portion of type β leaders is precisely what would be expected if these portions corresponded to the rapid first stage observed photographically for a leader passing down to a space-charge region in the air below the cloud, the fast step streamers involving exceptionally large and sudden changes of moment in each step. The absence of pulsations on the second portion of such records is in accord with the deduction from the general slope of the field change and from the photographic evidence that in this stage the leader process is slow and lowering a much smaller charge along its channel.

The oscillograph record of the *b* or return streamer process shows that the transient pulsation begins at the moment the return streamer starts from the ground, when the change of moment in the comparatively slow leader streamer is replaced by a much faster one due to the rapid passage of the leader charge to ground and to the creation of a positively charged stem to the upward-moving return streamer. The details of this and the step-streamer transient pulsations are now being studied with instruments of higher resolving power.

5. FIELD CHANGES DUE TO DISCHARGES WITHIN THE CLOUD

Oscillographic records have been obtained of many hundreds of lightning discharges which are known from direct observation to have taken place within the thundercloud itself. The distinguishing feature of these records lies in the absence of the sudden *b* change of field which is characteristic of a discharge to ground. The usual field change from a cloud discharge is a simple slow rise to a final maximum, this rise carrying superimposed pulsations of the same kind as those found for the stepped leader process and with the same time intervals between pulsations. Examples are given in fig. 17 (Plate 3). The explanation of this form of field change is presumably to be found in the observations made with the Boys camera on "air discharges" (Schonland, Malan and Collens 1935) which consist of a series of step streamers without a return streamer at their close. These air discharges are flashes which occur from the cloud to a space charge in the air, and it is legitimate to suppose that flashes which take place between different parts of the cloud have the same nature. This conclusion is confirmed by the oscillographic records which in general are identical in form with the leader portion of a type β first stroke.

The absence of a *b* portion with its powerful radiation component in these field changes suggests that cloud discharges are not usually such effective sources of distant atmospherics as discharges to ground, and that further study of atmospherics would enable those caused by cloud discharges to be distinguished from those caused by ground flashes.*

This work forms part of the lightning investigation sponsored by the South African Institute of Electrical Engineers whom we thank for financial support. We are very much indebted to the Victoria Falls and Transvaal Power Co. for the loan of the lorry used in the investigations. Our thanks are due to Dr D. J. Malan for frequent assistance in the field and to Professor Paine and his staff for many facilities.

To Mr R. A. Watson-Watt, Mr F. Lutkin and the late Mr J. F. Herd of the Radio Research Board of Great Britain we owe the opportunity of studying their technique and a great deal of helpful advice.

REFERENCES

Appleton, E. V. and Chapman, F. W. 1937 *Proc. Roy. Soc.* A, **158**, 1.
Collens, H. 1937 *Trans. S. Afr. Inst. Elect. Engrs*, **28**, 214.
Jehle, H. 1933 *Z. Phys.* **80**, 19.
Lutkin, F. 1937 *J. Sci. Instrum.* **14**, 209.
Malan, D. J. and Collens, H. 1937 *Proc. Roy. Soc.* A, **162**, 175.
Schonland, B. J. 1937 *Phil. Mag.* **23**, 503.
— 1938 *Proc. Roy. Soc.* A, **164**, 132.
Schonland, B. J. and Collens, H. 1934 *Proc. Roy. Soc.* A, **143**, 654.
Schonland, B. J., Malan, D. J. and Collens, H. 1935 *Proc. Roy. Soc.* A, **152**, 595.
Watson-Watt, R. A., Herd, J. F. and Lutkin, F. 1937 *Proc. Roy. Soc.* A, **162**, 267.

* This difference between "cloud" and "ground" atmospherics has been established in recent studies in this Institute. (B. S.)

X-Ray studies of the structure of hair, wool, and related fibres

IV. The molecular structure and elastic properties of the biological cells

By H. J. Woods

Textile Physics Research Laboratory, The University, Leeds

(*Communicated by Sir William Bragg, O.M., P.R.S.—Received 5 January* 1938)

[Plates 4, 5]

Introduction

Amongst the difficulties encountered in the study of the structure of hair not the least is its heterogeneity, whether considered crystallographically or histologically. From the former viewpoint progress has been made possible by assuming that there is no clear demarcation between "crystalline" and "amorphous" keratin, but rather that there is a continuous gradation of organization, so that the properties of "crystalline" keratin, which lies at one end of the scale, may be linked by a series of almost imperceptible changes with those of the keratin of any other degree of regularity. The X-ray photographs of the better organized parts have been regarded as signposts pointing the way to an understanding of what is happening, in the molecular sense, in the rest of the fibre substance.

Histologically, it is possible to define to a certain extent the part played by the various structural elements by comparing the properties of different types of hair. In this way we find that all hairs have certain basic similarities in their behaviour, and if we confine our attention to these we may at once eliminate, as being of no fundamental importance, such special histological features as the cuticle or the medulla, which occur in certain types of hair only, or can be removed from the specimen. There remains, as the seat of the characteristic properties of keratin fibres, only the cortex, which in all hairs consists of a mass of biological cells apparently fused together by an "intercellular medium". The present work is an attempt to discriminate between the cortical cells and the intercellular material in their relation to the elastic behaviour of the fibre.

In an earlier paper of this series (Astbury and Woods 1933)* it was shown how the normal extension range (up to twice the initial length of the fibre) can be divided into three steps, whose limits correspond roughly with the limiting extensions in dry air, cold water, and steam. The course of the intramolecular transformation as shown by X-rays is also roughly divided in the same way, for when a fibre is stretched by an amount not exceeding some 20 % there is very little change in the X-ray photograph, whilst further extension up to the limit in cold water is accompanied by a progressive replacement of the normal α-photograph by that of the β-form, so that at extensions of 60–70 % the original photograph has almost, if not entirely, disappeared. Whether there is an increase in the intensity of the β-photograph in the final stage of extension when the fibre is stretched in steam is difficult to say, but the provisional conclusion has been reached that it is chiefly within the intermediate range of extension that the best organized keratin transforms. The results of a number of experiments support the idea that the intermediate range is distinguished from the initial and final stages of the extension, for very often the rate at which an effect increases with increasing extension, whilst sensibly constant within each range, changes considerably in passing from one range to another. Sometimes only one of the limits is shown in this way; for instance, when the temporary set† of a fibre relaxed† in steam for a few minutes is plotted against the relaxation extension, it is found that there is a distinct lag in the development of the effect for extensions below 20 %; this is shown by the curve *b* in fig. 3.

Such observations can only mean that the fibre substance can be divided into three "phases" which transform successively in the three ranges of extension, and whose properties differ in certain respects; the phases, however, are to be thought of only as modifications of a fundamental keratin, and even then as modifications which are variable within themselves and between which there exist no clear-cut distinctions. The phases, moreover, do not function entirely independently, but restrict each other to such an extent that it is necessary to postulate that they act in parallel as well as in series. This state of affairs may be best visualized by supposing that K_1, the phase which transforms most easily, forms an effectively continuous matrix in which the others, K_2 (most "crystalline") and K_3, are embedded. An attempt was made (II) to apply this concept by identifying the combined K_2 and K_3 with the cortical cells, so that the intercellular medium would consist of K_1, and it was partly with the object of testing this suggestion that the present investigations were started.

---

* Referred to as (II) throughout this paper.

† When a fibre is held stretched at a constant extension in the presence of cold or hot water, steam, or alkali, its tension falls at a rate which depends on the temperature, etc., of its environment. When this *relaxation* takes place in cold water, the fibre still retains its power of reverting to its initial length when released in cold water; under other conditions of relaxation it may lose this power, wholly or in part, so that when released in cold water it exhibits a "*set*". "Set" imposed in this way may be diminished or entirely removed by increasing the temperature of the water in which the fibre is released. The residual extension when the fibre is allowed to contract freely in cold water is thus *temporary set*; the test for *permanent set* is to allow contraction to take place in steam or boiling water. A fibre may also be set at an extended length simply by drying; it is then said to have acquired a *cohesive set*, which may be removed by wetting the fibre. After certain relaxation treatments, hair acquires the property of contracting, under suitable conditions, to less than its original length; it is then said to be capable of *supercontracting*.

The points at issue reduce to two relatively simple questions: do the cells contain any crystalline keratin, and do they contain any K_1? The first is easily answered by investigating the cells themselves by X-ray methods, and it turns out that the cells from all the fibres studied (wool, whalebone, cow's horn, and human finger-nail) give, when oriented, X-ray photographs closely approximating to those of the fibres. The second question is more difficult, and can only be answered by a study of the elastic properties of the isolated cells. Since, however, they are too small to be manipulated singly, resort must be made to indirect experiments. In these investigations two methods have been used: in one, the fibres have been treated in various ways before disintegration, and the properties of the resultant cells studied; in the other, cells from normal fibres have been compounded into coherent films whose elastic behaviour was examined.

The results obtained leave no doubt that the properties of the fibre are, in the main, the same as those of the cells, which may be stretched, relaxed, set, and supercontracted in the same way as the fibres. In particular, the cells do stretch, during the initial extension of the fibre, almost as much as the fibre as a whole, so that it is clear that they do contain a proportion of the phase K_1. Whether the intercellular material may be classed with the cuticle and medulla as being of no basic importance from the point of view of the elastic behaviour is, however, doubtful; it seems more likely that it, too, has the characteristic keratin properties. It is even possible that the term "intercellular" is really a misnomer, that the material removed during the disintegration of the fibre into cells extends, too, inside the cells, which are only regions in this all-pervading medium reinforced, perhaps, by the presence of a greater proportion of the more highly organized keratin.

MATERIAL

The earlier, and to a certain extent preliminary, experiments were done with a sample of cells isolated by Waters (1932) and Gabriel (1932) by the action of the "Pink-rot organism" on New Zealand Romney wool. Later it was found necessary to prepare our own cells by tryptic digestion, following Burgess' procedure (1934). The great advantage of bacterial or enzyme retting lies in the moderate *p*H values used, so that the cells may be expected to be less seriously damaged than those which can be isolated by the older methods in which fibres were treated with strong acids or alkalis (Nathusius 1894). With the exceptions of the New Zealand Romney wool mentioned, and whalebone, which was easily disintegrated by treating the fibre with a dilute ammonia solution, each kind of hair used was retted by trypsin solutions buffered to *p*H 8·6. The fibres studied were Cotswold, Wensleydale, Lincoln, camel, and merino wools, cow's horn, and human finger-nail, as well as whalebone and New Zealand Romney wool. In every case the cells were purified by repeated washing in large volumes of distilled water, and care was taken to remove all buffer salts in those which had been retted by trypsin.

X-RAY PHOTOGRAPHS OF NORMAL CELLS

An account of the technique for obtaining fibre photographs of normal wool cells has already been given elsewhere (Woods 1936). Briefly, the procedure is based on the fact that elongated particles in an electric field tend to be oriented, if they are free to rotate, with their lengths parallel to the electric force (Pidduck 1925). The shape of wool cells is ideal for such an effect, since they are about 100μ long by a few microns thick, and when an alternating field is maintained across a suspension of the cells in a pool of water lying on a glass plate they are rapidly pulled into alignment and sediment out to form a layer in which they are approximately parallel to one another. The use of an alternating field is, of course, essential, on account of cataphoresis, which would cause a drift of the cells in a steady field. On drying, the oriented cells form a rather fragile tissue* which may be removed from the glass with a sharp razor.

X-ray photographs of tissues prepared in this way show very little dispersion when the beam is parallel to the plane in which the cells lie

---

* To avoid any possible confusion, it is well to state here that the term "tissue" is used in this paper to describe the filmy aggregate of cells prepared as described, and never in the biological sense.

(fig. 1 *a*, Plate 4), and even with the beam perpendicular to the plane the semi-angle of dispersion may be as small as 13°. Fig. 1 *b*, Plate 4, is a comparison, by the "quadrant photograph" method described in II, of a cell tissue (X-ray beam parallel to the plane of the tissue) with a bundle of parallel normal fibres. In this tissue the semi-angle of dispersion was about 20°, whereas in the fibre bundle it was about 9°. It will be seen in the figure that the intensity of the meridional group of spots (120), (020), (Astbury and Street 1931)* is considerably less in the cells than in the fibres, the estimated ratio being between 1/3 and 1/2. The specimens were adjusted to have approximately equal masses in the X-ray beam, and equal exposures were given to the two parts of the photograph, so that at first sight there would appear to be a discrepancy. A correction must be made, however, for the dispersion of the cells in the plane of the tissue. This may be done with sufficient accuracy by noting that the dispersion in the plane perpendicular to the X-ray beam is the same in the tissue and the fibre bundle, so that the intensities, for equal scattering powers, should be roughly inversely proportional to the dispersions given above. In view of the difficulties involved in any exact discussion of the intensities in such a case, the agreement must be regarded as satisfactory, and we may say that to a first approximation the cells diffract about as strongly as the fibres.

The success of the method in the case of wool cells suggests that other types of cell, not so greatly elongated, might also be at least partially oriented in an electric field. Whalebone and cow's horn, for instance, can be disintegrated into cells which are in the form of flat, moderately elongated disks.† Tissues of whalebone cells, oriented as described, were found to give fairly good fibre photographs with dispersions comparable with that in the fibre; and horn cells, being slightly less elongated but otherwise similar to those from whalebone, were oriented to about the same extent. Fig. 1 *c*, Plate 4, is an X-ray photograph of a tissue of horn cells; the dispersion, though considerably greater than that in the wool-cell tissues, compares favourably with that shown by many specimens of horn.

In the cells so far discussed the electrical orientation can be ascribed to the shape of the cells. From finger-nail, however, cells are obtained which are much more like flat circular disks, and in these it would appear that any orientation would be due rather to a difference in the dielectric constant

---

* Referred to as (I) throughout the paper.

† For photographs of whalebone, horn, and finger-nail cells, see Astbury and Sisson 1935 (referred to as (III) throughout this paper). The horn cells shown there seem to be exceptionally round; usually they are much more like whalebone cells.

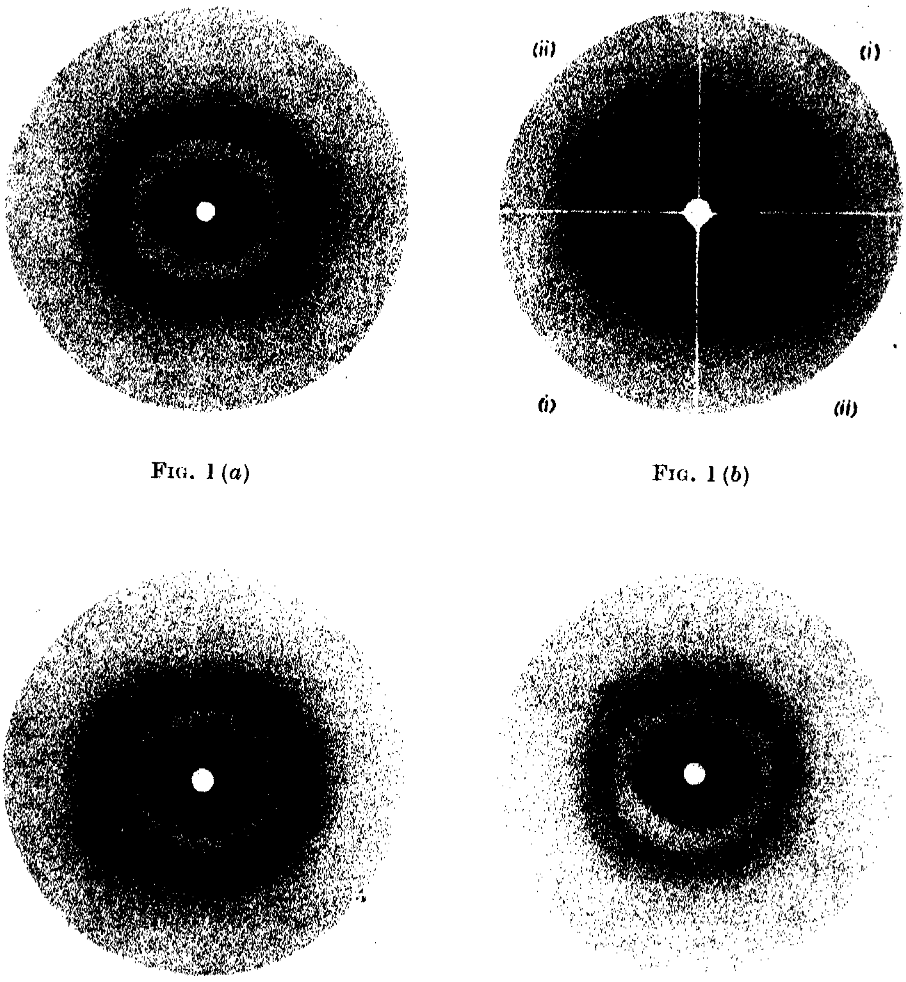


Fig. 1 (*a*)

Fig. 1 (*b*)

Fig. 1 (*c*)

Fig. 1 (*d*)

Fig. 1. X-ray photographs of cells from keratin fibres.

(*a*) Oriented cells from New Zealand Romney wool.

(*b*) Comparison of Wensleydale wool fibres (i), with oriented cells from Wensleydale wool (ii).

(*c*) Oriented cells of cow's horn.

(*d*) Oriented cells from steam-set Wensleydale wool.

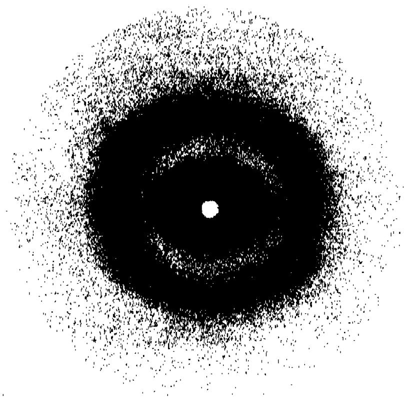


FIG. 1 (*e*)

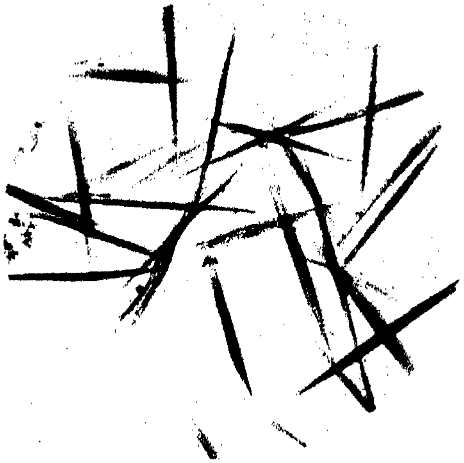


FIG. 2 (*a*)

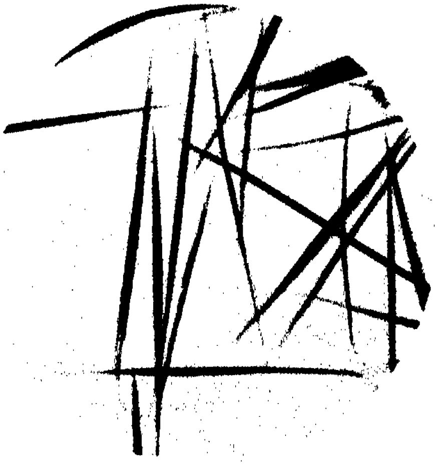


FIG. 2 (*b*)

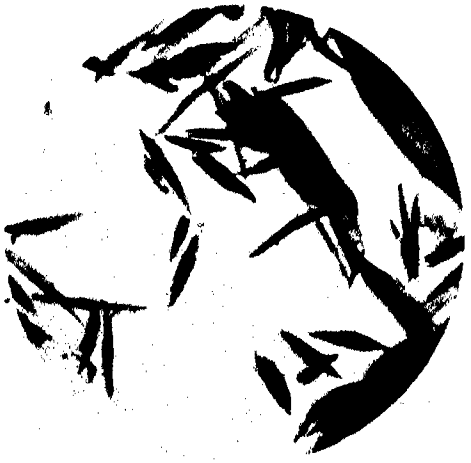


FIG. 2 (*c*)

FIG. 1. X-ray photographs of cells from keratin fibres.

(*e*) Coherent cell sheet stretched by 98 % in steam.

FIG. 2. Photomicrographs of wool cells (× 250).

(*a*) Normal Cotswold wool cells.

(*b*) Cells from steam-set Wensleydale wool (80 % extension).

(*c*) Supercontracted Cotswold wool cells (see text for details).

for fields parallel and perpendicular to the protein chains. When the test was made it was found that the cells did orient to a certain extent, and in the X-ray photographs, even when the beam was perpendicular to the tissue, it was possible to see the reflexions as arcs and not circles. That there was an effect due to the electrical anisotropy was supported by the observation that a cylinder of cow's horn about 1 cm. in diameter and 1·2 cm. long, cut so that the fibre direction was perpendicular to its axis, when hung by a silk fibre so that it could rotate about its axis between the charged plates of a small parallel-plate condenser, always tended to swing round until the fibre direction was parallel to the field. The orienting effect of the structure is thus in the same sense as that of the external shape, so long as the greatest length of the particles is parallel to the natural fibre axis. The two effects, however, are of quite different orders of magnitude in wool cells, and probably also in horn, since when elliptical disks of cow's horn with the fibre direction parallel to the minor axes were substituted for the cylinder in the above experiment, they oriented normally (with the fibre direction parallel to the field) only so long as the eccentricity was less than about 0·4. Except with particles which are very nearly circular disks or spheres, therefore, it is only the external shape which is of any importance in determining the direction and extent of the orientation.

MOLECULAR ORIENTATION IN PROTEIN FILMS

The experiment just described, where a cylinder of cow's horn tends to rotate until the direction of the main chains is parallel to an applied alternating electric field, may also be used as a test for determining the direction of the orientation of the chains in films of myosin or denatured proteins. Hitherto, the only unequivocal method has been to take an X-ray photograph, since optical tests may be misleading for various reasons.

A film of, say, denatured edestin, in which the chain bundles are in random orientation, as is shown by the X-ray photographs (Astbury and Dickinson 1935; Astbury, Dickinson and Bailey 1935), may be expected to be isotropic in its electrical behaviour. When a disk about 5 mm. in diameter and about 0·2 mm. thick was hung by a suspension through its centre perpendicular to its plane in an alternating electric field of about 100 V/cm. parallel to its plane, it was found to remain in equilibrium in all positions. Similar disks, cut from films which had been stretched by about 200 % in water and then dried, behaved in the same way as the keratin cylinder, that is, tended to rotate until the direction of stretching lay parallel to the applied field. It is clear, therefore, that the act of stretching, which orients the chain bundles in the direction of extension, also results in the development of electrical anisotropy.

Similar tests were applied to films of myosin. With myosin there is considerable difficulty in making a film in which the crystallites are arranged at random; accordingly, the electrical behaviour was found to vary considerably, some specimens cut from unstretched films showing an effect when hung in the field, and others behaving like the normal edestin films. An effect was always obtained with disks cut from stretched films, however, in accordance with expectations. In order to resolve the anomalies in the case of unstretched films, one of the electrically anisotropic specimens was therefore examined by X-rays, with the result that the orientation indicated by the electrical test was indeed found to be present, and the average direction of the chains, as shown by X-rays, coincided with the direction predicted. So far, owing to the difficulty of preparing stretched and unstretched films of the same uniform thickness, it has not been possible to obtain quantitative comparisons between different proteins or between films of the same protein at different extensions; the results obtained, however, show that the effects are of the same order in α-keratin and the other proteins, whether these are in the α-configuration (oriented myosin) or in the β-form (oriented edestin and highly extended myosin).

CELLS FROM STRETCHED FIBRES

According to the folded chain theory of the structure of α-keratin, the (approximate) numerical relation $2 \times 5{\cdot}1 = 3 \times 3{\cdot}4$, found between the longitudinal periodicities in α-keratin (5·1 A) and β-keratin (3·4 A) is a consequence of the elongation of the crystallites by 100 % during the transformation, so that three residues, which in the folded chain occupy a length of 5·1 A, have a length of 10·2 A in the straight chain. This relation corresponds to the observed extensibility of hair in steam, which is about 100 %, if we assume that at the limiting extension the whole of the fibre substance is elongated by the same amount. The histological heterogeneity of the cortex introduces an element of uncertainty in this connexion, however, since it is possible that the maximum extensibility of the fibre as a whole is not a measure of that of the intercellular material and the cells separately, but is only a mean value of their extensions at the breaking point of the fibre. Since the cells contain crystalline keratin, a more stringent test would be to compare the elongation due to the transformation with the extensibility of the isolated cells. If the latter can be measured accurately and should be found to differ greatly from 100 % the relation mentioned above would lose its significance. The possibility envisaged here, it should be noted, is something quite apart from any effect due to the phases, and it raises the question whether the molecular configuration in the intercellular material is, in fact, that of α-keratin, or whether the chains there are normally folded in some other way. A test for an effect of this sort would be to measure the extension of the cells for *high* fibre extensions; at low extensions complications might be introduced by the phases, if the effective proportions of these differ in the intercellular and cellular material (see below).

Previous studies have shown that "permanently set" keratin, obtained by steaming stretched fibres, retains its molecular β-configuration under a variety of subsequent treatments, but whether it would do so during digestion by trypsin was not known. Trial experiments with Cotswold wool fibres set in steam for 3 hr. at an extension of 80 % showed that the retting proceeds at a greatly enhanced rate, so that the cells themselves may be seriously attacked after a few days. Owing to the scale sheath retaining its form and holding the cells in bundles even after they were freed from one another, some difficulty was found in obtaining complete cells, continuation of the retting causing them to break up, at first into short lengths and ultimately into granules. In later experiments with Cotswold wool fibres which were relaxed in dilute caustic soda before retting a similar difficulty arose, and the solution proved to be to scrape the fibres after they had been retted for a short time; the scales were then found to come away from the more or less intact cortex, and the latter was returned to the trypsin to complete the disintegration.

In order to reduce the necessity for this treatment as much as possible, the present experiment was carried out with another kind of wool, Wensleydale, which has a smooth (and apparently thin) cuticle which broke up a good deal more easily than that of Cotswold wool. When Wensleydale fibres, after steaming at 80 % extension for 3½–4 hr., were treated with trypsin, cells were obtained in large numbers in less than a day, and no difficulty was experienced in collecting enough to make into tissues for X-ray examination. When seen under the microscope these cells appeared much more needle-like than normal cells, and there seemed to be no doubt that they were of the right length to account approximately for the fibre extension. Fig. 2 (Plate 5) shows how they compare with the normal cells.* Owing to the large variation in the lengths of the latter, a cursory examination is not entirely satisfactory, and an attempt was made to compare the mean lengths of the normal and elongated cells, with the results shown in Table I. It will be seen that the stretched cells are about 75 % longer on the average than the normal cells, and although too much reliance should not be placed on this actual value, owing to the large standard errors of the means, the figures are sufficient to show that the cells in fibres stretched to 80 % extension are extended by approximately the same amount (see below).

---

* See footnote, p. 89.

Tissues made from the elongated cells by the usual device were not so well oriented as those made from normal cells, owing to a pronounced tendency which the stretched cells had to flocculate. The X-ray photographs of these tissues showed all the features of the ordinary photograph of β-keratin, except that the "spreading" of certain spots along the hyperbolae was not apparent. This might easily have been due to the angular dispersion in the tissues. Fig. 1*d*, Plate 4, shows an X-ray photograph of Wensleydale wool cells in the β-form prepared by the method described.

TABLE I. COMPARISON OF THE LENGTHS OF NORMAL AND ELONGATED CELLS (WENSLEYDALE WOOL; ELONGATED CELLS FROM FIBRES AT 80 % EXTENSION)

	No. of cells	Mean length μ	Standard deviation μ
Normal	208	106	22
Elongated	200	185	36

Cells measured dry (room humidity).

TABLE II. RECOVERY OF ELONGATED CELLS IN BOILING WATER

Time of boiling hr.	No. of cells	Mean length μ	Standard deviation μ
0	200	185	36
1	87	164	43
3½	98	153	39
7	102	148	30

Cells measured in water, except those for 0 hr.

That the cells themselves take part in the elastic phenomena is thus clear; it appears also that the temporary set developed in fibres when stretched and steamed is not due to the intercellular material, but is a property possessed by the cells as units. They probably also, therefore, are capable of being permanently set, i.e. the recovery in boiling water of the elongated cells should be only partial. In accordance with this expectation it was found that after boiling for 7 hr. they still remained about 41 % longer than the normal cells. For fibres steamed for 4 hr. at 80 % extension we should expect about 45 % permanent set (this value is taken from II, corrected for swelling, since the cells were measured in water). As a first approximation, therefore, it is possible to take the behaviour of the cells in this experiment as identical, under the same conditions, with that of fibres; certainly there is no evidence that the cells set more than the fibres.

The small contraction of the steam-set cells during the retting process (as judged from their average length when isolated) is important in another respect, because it means that from a study of the lengths of the cells from fibres set at various extensions some information may be forthcoming about the way in which extension is realized in the intact fibre; whether, that is, at low extensions the intercellular material is responsible for the whole of the stretching, or whether the cells also make a contribution. This seems to be the most direct test of the suggestion put forward in II, for if it turns out that the cells themselves stretch at low extensions, then they must contain some K_1. Experiments were therefore made to determine the relation between the extension of steam-set fibres (in each case the steaming was for 4 hr.) and the mean length of the cells obtained from them. Table III gives the figures for cells from Cotswold wool fibres steamed at extensions up to 53 %; the cells were measured wet (in water), and for comparison the wet length of normal Cotswold wool cells is also given.

TABLE III. LENGTHS OF COTSWOLD WOOL CELLS FROM FIBRES SET IN STEAM AT VARIOUS EXTENSIONS

Extension of fibres %	No. of cells	Mean length μ	Standard deviation μ	Extension of cells %
10	224	114	25·4	5
15	202	118	25·4	8
19·5	130	121	26·4	11
20	246	120	26·4	10
30·5	203	131·5	27·0	21
40	216	144·5	27·7	33
53	203	155	30·4	42

All measured in water. Normal length of wet cells, 109 μ.

When the results are plotted as in fig. 3 *a*, it appears that there is a lag in the increase in length of the cells with increasing extension of the fibres below 20 %, an effect which recalls the lag in the development of temporary set in the same range for fibres; the latter is shown by the other curve in the figure. The value of the discrepancy, 9·5 % at 20 % fibre extension, is rather greater for the cells than for fibres steamed for corresponding times, as was shown by measurements of the temporary set (in the buffer solution at the retting temperature) of fibres steamed for 4 hr. at various extensions. The actual recovery of a fibre after steaming at an extension of 20 % was 2·8 %, which leaves unaccounted for a deficiency of about 6½ % in the average cell length at 20 % extension.

It is clear, however, that since the cells stretch at low extensions of the fibre they must definitely contain some K_1. Two explanations are then possible for the shape of the curve in fig. 3. The first is that here again we

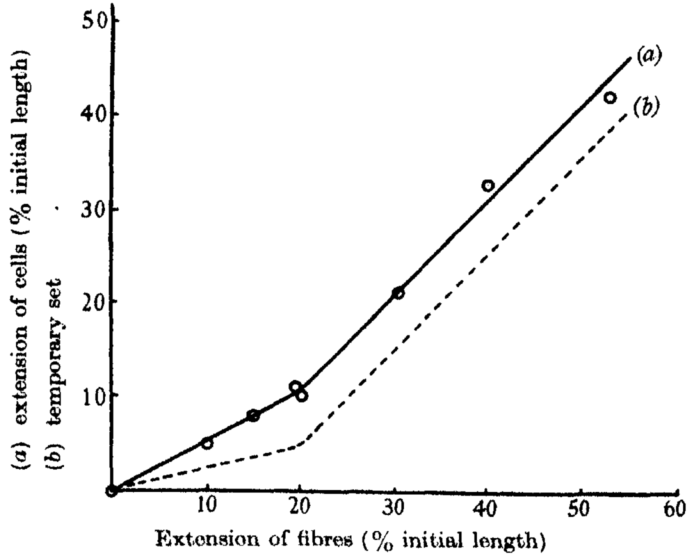


FIG. 3. (a) The relation between the mean length of cells from Cotswold wool fibres set in steam for 4 hr. and the extension at which steaming took place. (b) The relation between the temporary set of human hair fibres steamed for 2 min. and the extension at which steaming took place.

have evidence of the smaller setting power of the K_1 phase. The set induced in K_1 by steaming for 4 hr., though almost complete when the fibre is released in cold water, is only partial when the fibre is allowed to contract in steam. If we suppose that trypsin is able to break those linkages which are responsible for temporary set in K_1, but not those in K_2 and K_3, then its contractile effect on K_1 will be similar to that of steam. The phenomenon measured in the present experiment would then be something more closely related to the *permanent* set of the fibre than to the extension at which steaming took place. If this view is adopted the cells may be regarded, with respect to the phase K_1, as representative of the fibre as a whole.

It is possible, however, that trypsin has not this power of releasing temporary set. Certainly, as the experiment at 80 % extension shows, the recovery of the cells during retting is small compared with the residual temporary set which can be removed by boiling the cells, so that retting is not effective in destroying the linkages responsible for temporary set in K_2 and K_3. From this point of view, the discontinuity in the slope of the curve is due to the extension of the cells lagging behind that of the fibre. This would indicate that K_1 contributes less to the elongation of the cells than it does to that of the fibre as a whole.

Which of these two explanations is correct is difficult to decide. A possible test depends on the slope of the curve for the higher extensions, for according to the first hypothesis it should be not greater than 1, which is the value it would have if the recovery during retting is confined to K_1. If, however, the other suggestion holds, the slope should be greater, for the cells would have to regain the ground lost during the initial part of the extension. If they do so by the time the fibre has stretched to 100 % the slope should be about 1·1. The calculated value for the line of best fit for the last five observations of Table III is 0·99, but the reliability of this is hardly great enough to allow a definite conclusion to be drawn. If anything, the balance of evidence is in favour of the first hypothesis; but against this must be placed the measurements for Wensleydale cells given in Table I. The apparent recovery of these cells, which was about 5 %, is really rather an over-estimate, for if the measurements are reduced to wet lengths, as are those in Table III, the measure of agreement between cell and fibre extensions becomes even closer, on account of the greater percentage swelling of the elongated cells.

Since the cells contain some K_1, the simple picture of the histological structure of the fibre put forward in (II) is clearly invalid. A better approximation to the truth is probably that the cells are regions in a continuous phase K_1 reinforced to a certain extent by the occurrence of the other two phases. The latter may possibly be associated with the fibrils (Nathusius 1894) which can be obtained by further treatment of the cells by alkalis or acids, and whose presence can be discerned even in cells which have been in trypsin for long periods. It seems to be well established that the process of digestion by trypsin does actually cause the cells to get thinner if it is prolonged (Burgess 1934), although this is not accompanied by any very great shortening, since the mean length of normal cells seems to be constant for a given kind of wool, even when no particular precautions are taken to ensure uniform retting. This would be the case if the length of a cell is determined by the fibrillar content; the diameter would depend more on the amount of interfibrillar material. There is also reason to believe that the fibrils give the keratin X-ray photograph, since Rudall (1936), in a study of the effect of compression on horn and porcupine quill, finds that in certain experiments where the photograph is a composite one of two α-photographs at an angle, the fibrils tend to form a sort of net with this angle. He deduces that they are "crystalline", and suggests that K_1 is the interfibrillar material. At the moment, therefore, it appears that this idea may ultimately prove to be correct.

SUPERCONTRACTION

Further evidence that the cells behave elastically in a similar way to the fibre as a whole is afforded by a study of supercontraction in the cells. Two possible methods are available for isolating cells in the supercontracted state; retting supercontracted fibres is the obvious way, but unfortunately all attempts to do so failed to give cells which were in any respect satisfactory. Apparently the effect of supercontraction by the action of steam on relaxed fibres is to modify the intercellular material to such an extent that it is no longer preferentially attacked by trypsin. The fibres are certainly more resistant than after relaxation without supercontraction, as is perhaps to be expected, since, as pointed out in (II), setting effects must necessarily accompany contraction in steam; so that failure to obtain good cells is not due to an increased susceptibility of the cells, but rather to a stabilizing of the intercellular material. It was thought that this might be a consequence of the greater degree of folding of the protein chains in the supercontracted fibres, and that pepsin might therefore be more efficient than trypsin in removing the "cement", but it was found on trial that digestion in pepsin was, if anything, less successful than in trypsin.

The alternative process was much more successful. It was based on the fact that a fibre could be relaxed sufficiently to induce a considerable capacity for supercontraction without damaging the cellular material so much that it would not remain intact during the retting process. The actual details of a typical experiment were as follows: Cotswold wool fibres were treated at 50 % extension for 10 min. with a 1 % aqueous caustic soda solution; they were then washed thoroughly to remove the alkali, and put into the retting solution. After a few hours they were transferred to a glass plate to form a bundle, which was scraped with a moderately sharp glass edge (the end of an ordinary microscope slide was found to be suitable) in order to remove as many of the scales as possible. The fibres were then returned to the retting bath, and in a few hours more the cells separated quite well. The mean length of the cells obtained in this way was about the same as the normal, so that the temporary set shown by fibres relaxed in caustic soda was absent. This is not surprising, since the fibres and cells had been standing in a warm alkaline solution, and we know that temporary set of this kind can be removed by treatment with dilute caustic soda. After washing, the cells were boiled in distilled water for about an hour, and this caused them to shorten to about half the normal length. Table IV gives the results of measurements of their lengths, and fig. 2 *c*, Plate 5, is a photomicrograph of some of them. The three photographs in this figure show how great the changes in the form of the cells may be under various treatments.*

TABLE IV. COMPARISON OF NORMAL, RELAXED, AND SUPERCONTRACTED CELLS (COTSWOLD WOOL)

Treatment	No. of cells	Mean length μ	Standard deviation μ
Normal	320	105	18
Relaxed	72	97	18
Supercontracted	95	53	8

All measured at room humidity.

During the treatment with boiling water the cells flocculated so badly that any attempt to form an oriented tissue was out of the question. An X-ray photograph of some of the cells which were allowed to sediment on a glass plate without being aligned (this gave a tissue with the cells approximately oriented with their lengths parallel to a plane) showed that the crystalline keratin was in the β-form, but that it was not completely disoriented in the cell. In this respect the cells behave in the same way as the fibres.

COHERENT FILMS FROM WOOL CELLS

The experiments described above are sufficient to tell us a good deal about the way in which a fibre stretches, so long as we assume that when fibres are treated in various ways before retting the material which is removed by the trypsin is the same. Actually, there is no guarantee that this condition is rigorously fulfilled, for it is quite clear that some treatments (notably, relaxation in dilute caustic soda) are difficult to control so that the fibres will afterwards disintegrate into recognizable cells. On this account it was thought worth while to search for a method of stretching the normal cells. In the absence of a micro-technique capable of dealing with the cells singly, the only possibility is to make them cohere so strongly that they may be stretched in bulk, but to do this by embedding them in an adhesive matrix proved quite impossible, owing to the inextensibility of strong adhesives such as cellulose, and the weakness of extensible materials such as rubber or the proteins.

---

* Although the elongated cells shown here are from Wensleydale wool, whilst the others are from Cotswold wool, the normal cells of the two fibres are so similar that it is unnecessary to show both.

One experiment, however, should be possible without stretching the cells directly—the transformation from α- to β-keratin by the application of lateral pressure in steam (III)—and in order to show that this could be brought about in the isolated cells, a bundle of the oriented tissues was pressed between glass plates in steam for several hours with the rather surprising result that it was transformed into a transparent and exceedingly tough film, recalling, in appearance, a thin sheet of tortoise-shell. An idea of the change effected in this way may be gathered from the fact that the original tissues were too fragile to handle except very gently with forceps, and were quite opaque except when made extremely thin; the pressed films, however, were (in strips about 2 mm. wide and 100μ thick) too strong to be broken by gripping with the fingers and pulling, and were transparent enough to read through.

Properly made sheets of this sort not only remained coherent when placed in cold water, steam, or dilute caustic soda, but also had an extensibility comparable with that of the fibres (see below). It is remarkable that such coherence can be developed simply by the action of steam on the strained keratin, and it seems probable that a molecular linking up of neighbouring cells takes place similar to that between the polypeptide chains of the grid during permanent set (II). The suggestion is made more plausible by the fact that the development of resistance to steam or alkali requires *prolonged steaming* of the pressed tissues; pressure in water at ordinary temperatures merely results in a sort of cardboard which disintegrates when wetted, whilst insufficient steaming under pressure gives a transparent sheet which is stable in cold water but not in steam. These effects seem to be directly comparable with the phenomena of cohesive and temporary set in fibres (II). Cells which had been dyed after retting showed a great reluctance to cohere, so that apparently the basic-NH_2 groups in the keratin are involved in some way, just as they are in permanent set (Woods 1933; Speakman 1933, 1936; Astbury and Dawson 1938).

The degree of compression used in forming these sheets was less than that required to give a clear indication of the presence of β-keratin in the X-ray photograph; but examination of more drastically treated specimens showed that continued pressure resulted in the transformation. The photographs obtained, however, were never good, because the procedure involved a transformation at constant length (in contrast to that used in (III), where the effect of pressure was aided by the ability of the specimen to expand longitudinally), which could of course be realized only at the expense of crystallite orientation; complete disappearance of the α-photograph was never obtained, either. The possibility of stretching still remained, nevertheless, and after some practice it was found possible to make sheets which were extensible in cold water. At the highest extensions thus attained (rather less than 50 %) the X-ray photograph was still composite, although the amount of α left was not very great. In this respect, and also in that there seemed to be little change in the photograph at extensions less than 20 %, the films are very similar to fibres, so that the results support the conclusions drawn from the retting experiments. Although, as mentioned above, there is a loss of crystallite orientation in the pressed sheets, it must not be concluded that the disorientation was enough to account for any considerable fraction of the extensibility; for the most part the extension must be of the same nature as in fibres.

It was now clear that, just as in fibres, the complete β-photograph could be realized only at high extensions, which can be reached only in some more efficient relaxing agent than cold water. That the sheets could be stretched in steam seemed at first too much to expect, but trial showed that good specimens had an extensibility in steam (or dilute caustic soda) of the order of 100 %. At the higher extensions the β-photograph became more perfect, as shown in fig. 1 *e*, Plate 5, a photograph of a specimen stretched by 98 % in steam. All the features of the normal β-photograph are present, including the spreading of certain spots along the hyperbolae.

ELASTIC PROPERTIES OF THE PRESSED CELL-SHEETS (COTSWOLD WOOL)

(*a*) *Extensibility and long-range elasticity.* In cold water the highest extension realized was 48 %; elastic recovery in cold water was always complete (within 1 %) from extensions of this order. In steam or dilute caustic soda the extensibility was about twice that in cold water. Fig. 4 shows time/extension curves for specimens stretched under constant load in steam and in 2 % caustic soda (cf. (II), fig. 2); a limiting extension in the neighbourhood of 100 % is indicated by the shape of the curves. Specimens stretched to high extensions in the alkali, and then allowed to recover in the same solution, had residual extensions (wet) of about 10 %, but when washed and dried they contracted still further, to the initial length, at least, and in some instances to about 10 % less than the initial length. In corresponding experiments with Cotswold wool fibres, the final supercontraction realized is in the neighbourhood of 20 %. Both in extensibility and power of elastic recovery, therefore, there is no significant difference between the cell-sheets and fibres.

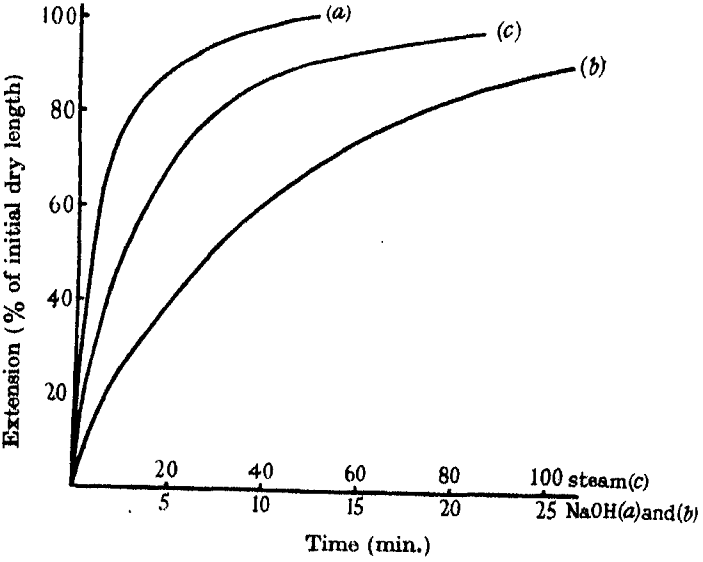


FIG. 4. Time/extension curves under constant load for pressed cell sheet in 2 % aqueous caustic soda (*a*) and (*b*), and in steam (*c*).

(*b*) *Load/extension curves.* In fig. 5 are shown load/extension curves for cell sheets, cow's horn, and human hair, under similar conditions of stretching (in cold water, at about 5 % per minute). The curve for the cells is almost exactly the same as that for a wool fibre which has been "derestricted" by repeated extension in cold water (e.g. the one on the left of (II), fig. 5); retting therefore appears to remove to a certain extent the restrictions between the phases. The relatively great effect of changing the rate of extension is shown by the broken curve, for which the entire stretching took only 90 sec.

(*c*) ***Relaxation and set.*** Internal breakdown is shown by the fact that repeated extension moves the load/extension curve to the left, just as in normal fibres. A capacity for temporary set is also possessed by the sheets, recovery in cold water after extension in steam being only partial. One specimen, after steaming for 30 min. at 48 % extension, recovered in cold water to 22 %; with normal fibres the corresponding recovery would have been only a few per cent, so that set takes longer to develop in the sheets. That there is a lag in the setting below 20 % is also possible, since one specimen, steamed for 30 min. at 17·5 % extension, showed a temporary set of only 4·3 %. In its capacity for permanent set also the sheet is inferior to fibres, since two strips steamed at 50 % extension for 6 hr. and then released in boiling water recovered to extensions of 21 and 17 %, respectively; normal Cotswold wool fibres similarly treated give about 28 %.

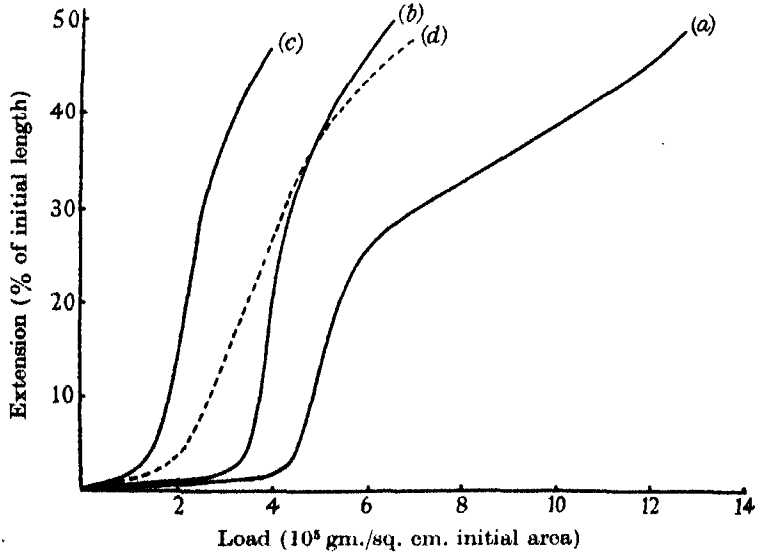


FIG. 5. Load/extension curves. (*a*) Human hair; (*b*) cow's horn; (*c*) pressed cell-sheet; all stretched at the rate of 5 % per minute in cold water. (*d*) Pressed cell sheet stretched in cold water as quickly as possible.

(*d*) ***Supercontraction.*** Sheets relaxed in 2 % caustic soda at extensions of 90 and 101 % and then washed, contracted in steam to 30·5 and 33 %, respectively, less than their original lengths. Normal fibres, under similar conditions of relaxation, supercontract by 50 %. The difference is not surprising in view of the treatment of the cells; the pressing in steam necessary to form coherent sheets would tend to decrease the supercontracting power. The results, however, are not inconsistent with the idea that part of K_1 has been removed, leaving K_2 in a proportion rather greater than normal; for we know that some at least of the crystalline keratin does not shorten during supercontraction, but actually lengthens* (II).

The sheets, therefore, show all of the characteristic properties of keratin in the form of fibres, and in the cases where measured values do not agree with those from fibres the discrepancies are only of the order we should expect in passing from one type of hair to another (e.g. the maximum supercontraction attained after relaxation in caustic soda varies considerably; it is about 50 % for Cotswold wool, and about 30 % only for human hair).

MISCELLANEOUS

In the course of these investigations a few other points have received preliminary attention. It was noticed that in washing the cells from wool fibres the cuticle scales, which are in the form of thin plates, tended to float on the surface of the water as the cell suspension was poured from one vessel to another. This made it possible to effect a partial separation of the two kinds of cells; and by starting with the mixture obtained by scraping incompletely retted fibres it was possible to obtain samples of the scales relatively free from cortical cells. So far, complete separation has not been attained, but X-ray photographs have been taken of tissues in which the estimated proportion of scales was about 80 %; they indicate that the degree of crystallinity and alignment of crystallites in the scale cells is poor compared with that in the cortex. This is in agreement with the optical properties, since a cursory examination suggests that the scales have a very small birefringence, and would also account for the observation that in highly supercontracted Cotswold wool fibres the scales appear to be traversed by lateral wrinkles, as if they cannot shorten longitudinally but must buckle as the cortex contracts. Since supercontraction in a randomly dispersed system could take place only at the expense of a change in density, and the measurements which we have made of the volume of a fibre before and after supercontraction show that there is little such change, any angular dispersion of the crystallites in the scales would result in a decreased power of supercontraction.

The proportion of intercellular material in the fibre is, as indicated elsewhere, probably an ill-defined quantity, and can hardly be expected to be measurable with any great accuracy. Thus, Stakheyewa-Kaverznewa and Gavrilov (1936) report a loss in weight of 10 % during the retting of an unnamed wool, whilst our own measurements show a considerably greater loss, $19 \pm 3\%$. The loss in weight is without doubt a function of the time of treatment, since the cells themselves are attacked in time. In this connexion the experiments of Haller and Holl (1936) and some of our own are relevant. The former find that when wool is treated with ammonia a water-soluble protein extract is obtained which they call lanain; and the fibres are broken up into cells, which indicates that lanain is a disintegration product of the intercellular material. In corresponding experiments with cells, we have obtained a similar extract, and it seems likely that the precursor of lanain also exists inside the cells. If this is so, it suggests that the extract is really formed by hydrolysis of the keratin constituting the phase K_1, and since prolonged treatment with ammonia causes the cells to break up into fibrils, a study of these seems to be the next step towards locating the crystalline phase.

---

* This point has recently been discussed at some length (Astbury and Woods 1937 *a*, *b*).

Further work is now in progress on the points raised in this section, and will be reported on in due course.

The investigations described in this paper have been made possible by the generosity of the Worshipful Company of Clothworkers, to whom the writer wishes to acknowledge his indebtedness. He also wishes to thank Dr W. T. Astbury for help given in numerous discussions, and Dr K. M. Rudall for advice on retting procedures and other matters.

SUMMARY

From a study of the cells isolated from various types of animal hair it is concluded that they contain crystalline keratin, and when the cells are long, tissues can be built up to give X-ray photographs closely comparable with those of the fibres. Cells obtained by retting fibres previously set in steam are found to be stretched correspondingly, although their measured lengths after retting are found to be somewhat shorter than would be expected. This lag is associated more with the early stages of extension, indicating that the elastic phase K_1 extends inside the cells, as well as between them. The X-ray photograph of elongated cells from fibres set at high extensions is the ordinary one of β-keratin. The elongated cells only recover partially when they are boiled in water, and the permanent set remaining is of the same order as that which the fibres would have shown. Permanent set is thus also a property of the cells themselves.

Cotswold wool fibres which have been relaxed in dilute caustic soda can also be disintegrated into cells, which recover during the process to rather less than the normal length. These relaxed cells can be supercontracted

still further by boiling water. In this way cells of about half the normal length can be obtained; their crystalline part shows itself in the β-form, just as does that of highly supercontracted fibres.

By the combined action of lateral pressure and steam the cells from normal fibres may be compounded into coherent transparent sheets which are elastic in cold water for extensions up to about 50 % and are extensible in steam or caustic soda by twice this amount. These sheets can be relaxed, set, and supercontracted, in the same way as fibres, and their tensile strength in water is about 25 % of that of fibres. It is suggested that the development of coherence during the treatment in steam is of a similar type to the development of permanent set in a stretched fibre.

Since such differences as exist between the properties of the cells and those of the fibres are only of the same order as those which occur between various types of hairs, it is clear that the behaviour of the fibres is for the most part due to the cells. The histological heterogeneity of hairs can affect only the details of their behaviour.

Keratin, stretched denatured edestin, and stretched myosin are all anisotropic in their electrical properties, having different dielectric constants for fields parallel and perpendicular to the protein chains. By hanging disks of these proteins in an alternating electric field it is therefore possible to determine the direction of the chains.

REFERENCES

Astbury and Dawson 1938 *J. Soc. Dy. Col., Bradford*, **54**, 6.
Astbury and Dickinson 1935 *Nature, Lond.*, **135**, 75.
Astbury, Dickinson and Bailey 1935 *Biochem. J.* **29**, 2351.
Astbury and Sisson 1935 *Proc. Roy. Soc.* A, **150**, 533.
Astbury and Street 1931 *Philos. Trans.* A, **230**, 75.
Astbury and Woods 1933 *Philos. Trans.* A, **232**, 333.
— — 1937 *a* *J. Text. Inst., Manchr*, **28**, 394.
— — 1937 *b* *Chem. and Ind.* **56**, 1076.
Burgess 1934 *J. Text. Inst., Manchr*, **25**, T 391.
Gabriel 1932 *J. Text. Inst., Manchr*, **23**, T 171.
Haller and Holl 1936 *Kolloidzschr.* **75**, 212.
Nathusius 1894 *Arch. Mikr. Anat.* **43**, 148.
Pidduck 1925 "A treatise on electricity", p. 66, 2nd ed.
Rudall 1936 Ph.D. Thesis, Leeds University.
Speakman 1933 *J. Soc. Dy. Col., Bradford*, **49**, 180.
— 1936 *J. Soc. Dy. Col., Bradford*, **52**, 335.
Stakheyewa-Kaverznewa and Gavrilov 1936 *Bull. Soc. Chim. Fr.* (Memoirs), **4**, 647.
Waters 1932 *Nature, Lond.*, **129**, 464.
Woods 1933 *Nature, Lond.*, **132**, 709.
— 1936 *Proc. Leeds Phil. Soc., Leeds*, **3**, 132.

The development of the spark discharge

By T. E. Allibone, D.Sc., Ph.D. and J. M. Meek, M.Eng.

(*Communicated by S. R. Milner, F.R.S.—Received* 31 *January* 1938)

[Plates 6–11]

1. Introduction

The development of the short electric spark has been extensively studied by various scientists who used the Kerr electro-optical shutter, which enables the spark to be photographed at short intervals after its initiation. The longest sparks so studied have been of the order of a few centimetres, and the gap has usually been between parallel plate electrodes. The spark has been shown to start sometimes at the electrodes, and sometimes in the mid-gap region. The cloud chamber has also been used by others to study the pre-discharge and the spark between dissimilar electrodes and between parallel plate electrodes: between dissimilar electrodes the discharge generally starts from the smaller electrode: between parallel plate electrodes the same results were obtained as had been found by the Kerr cell technique. The rotating camera has been used by the authors to photograph the long electric spark and preliminary results have already been communicated (Allibone and Meek 1937; Allibone 1938). The present paper extends the results given in the two preceding papers.

2. Apparatus

For the study of the long electric spark the impulse-voltage generator is the most convenient piece of apparatus for producing the high voltages required, and the time of application of the voltage to the spark gap may be easily measured by the cathode-ray oscillograph. The particular impulse generator used in this series of experiments has been described elsewhere (Edwards and Scoles 1937). Essentially it consists of twelve condensers, having a series capacitance of 0·01 μF, and a maximum output voltage of 2000 kV. The generator was connected to the spark gap through a series resistance comprising a bakelite tube containing water, the resistivity of which could be altered at will. The values of resistance employed were always sufficient to render the circuit aperiodic. For convenience one terminal of the impulse generator and one electrode of the spark gap were earthed. The length of spark investigated varied from 25 to 200 cm., and a variety of electrode arrangements have been used, such as point-point, point-plane, sphere-point, sphere-sphere. The point electrode consisted of a brass rod 1 cm. in diameter having a sharp tip tapered to a cone of 30°: the plane electrode consisted of a large sheet of steel, or of a trough of water: spheres of 12·5, 25 and 100 cm. diameter were employed. The high-voltage electrode was suspended vertically above the plane electrode.

To record the voltage applied to the electrodes connexion was first made from the high-voltage electrode through a small series resistance to a capacitance potential divider (35 $\mu\mu$F) connected to a high-speed cathode-ray oscillograph, the function of the resistance being to make this subsidiary circuit aperiodic. It was found that the development of the discharge was considerably affected by the presence of this parallel capacitance, so the divider was replaced by an antenna of length sufficient to pick up a suitable voltage for direct application to the oscillograph. The current in the spark was measured by earthing the lower electrode through a small non-inductive resistance and recording on the oscillograph the voltage so developed across the resistance. The minimum spark-over voltage has been employed for the majority of the discharges.

The spark was photographed by means of a rotating camera in which the film moved at a speed of 41·7 m./sec. at right angles to the image of the spark. The camera consisted of a brass drum 26·7 cm. diameter mounted on a horizontal shaft driven by a 3000 r.p.m. induction motor. The drum could be traversed 9 cm. along its axis, thus enabling a succession of sparks to be photographed and afterwards identified: this proved of great value when synchronized records were being taken on the camera and on the oscillograph. On the drum was fixed a 36 × 6 in. "Ilford Oscillographic Film Type F", and a special casing allowed the drum to be loaded in the dark room and then carried in the case to the camera chamber. The lens used was a quartz ultra-violet flint glass lens of aperture 3 cm. and focal length 12 cm., and light from the spark was reflected into the lens by a front surface aluminized mirror. Under the best conditions of focus and spark formation a resolving power of a $\frac{1}{2}$ μsec. can be claimed with this camera.

3. General description of records

All the photographs taken show that the spark consists of two or more components. In the simplest case a discharge is propagated from the high-voltage electrode to the earthed electrode and is then succeeded by a return discharge from the earthed electrode to the high-voltage electrode. Thus the formation of the spark discharge is generally analogous to that of the lightning flash, and the nomenclature developed by Schonland (Schonland and Collens 1933, 1934; Schonland, Malan and Collens 1935) to describe the lightning flash will be adopted in this paper. In an earlier publication (Allibone and Schonland 1934) the "leader-stroke" to the spark was only photographed for a fraction of the interelectrode spacing, but in the series of records described the leader stroke traverses the whole gap, and the return stroke or main stroke only occurs after the leader or leaders have fully developed. In the lightning discharge the Boys camera shows that the main stroke has a velocity of propagation about 100 times greater than the effective velocity of the stepped leader (Schonland, Malan and Collens 1935): a similar relationship exists for the leader and main-stroke velocities in the long spark discharge. It has not been possible to employ two lenses as in the Boys camera for the determination of the speed of propagation of the main stroke, but a lower limit to this speed can be determined by the following simple process: a number of rotating camera photographs were taken between two point electrodes, and on the same section of film a number of photographs of the spark were taken with the drum stationary. Many of the stationary records were sufficiently close to the rotating camera records to enable accurate measurements to be taken of the spacing between the stationary photographs and the main strokes of the rotating photographs. Such spacings at the high-voltage electrode were equal to those at the earthed electrode to within the limit of accuracy of measurement. Thus for a 150 cm. gap the speed of the main stroke could not be less than 150 cm. per ½ μsec., or 3×10^8 cm./sec. (corresponding to $3{\cdot}4 \times 10^8$ cm./sec. for the three-dimensional track velocity). The probability is that the speed is greatly in excess of this, as the oscillograms of voltage wave form applied to the gap show an almost instantaneous collapse of voltage. It is doubtful therefore whether a two-lens camera would be capable of determining the speed of the main stroke for gaps of the order of a few metres.

Due to the relatively high speed of the main stroke the speed of the leader stroke can be determined by direct measurement of the leader/main stroke spacing on the rotating camera records. It is obvious that in times of the order of 10–100 μsec. there is negligible movement of the air in the gap so that a separation of the leader/main stroke on the records is a separation in time and not in space. The projected track length is on the average 7 % greater than the interelectrode spacing, so that a further 7 % increase gives the approximate three-dimensional track, i.e. the three-dimensional track is 1·14 times the interelectrode spacing.

As there is such a very marked difference between the discharge when the polarity of the high-voltage electrode is changed, detailed discussion of the results will be separated into two categories, positive and negative discharges.

In all the illustrations, the high-voltage electrode is at the top and the earthed electrode is at the bottom, the leader stroke appears at the right and the return or main stroke appears at the left. The horizontal streaks extending to the left from each electrode in the spark photographs are caused by electrode incandescence. Other streaks across the photographs are due to discharges photographed on a neighbouring section of film. The oscillograms given in the text are printed with their time axes from right to left to correspond to the photographs of the sparks: a zero line is given for reference, and when a timing wave is shown this has a natural period of 8 μsec. The spacing, d, between the electrodes, the circuit resistance R, and, where relevant, the height, p, of the earthed point above the earthed plane are given on each photograph.

4. THE POSITIVE DISCHARGE

(i) *High-voltage positive-point electrode: plane earthed electrode*

Typical records of this type of discharge are shown in fig. 1, Plate 6. The distinctive feature is that the leader stroke starting from the anode traverses the whole gap to earth, branching downwards all the way: the main stroke follows in detail the tortuous path of the leader stroke. The leader appears as a sharply defined channel of width not exceeding 0·02 mm. on the photographic negatives indicating that the duration of high luminosity at any one point on the leader cannot be more than 0·5 μsec. It will be shown later that the velocities of leaders are approximately 2×10^6 cm./sec., so that the length of the highly luminous section of the leader stroke at this speed cannot be more than 1 cm. As the leader approaches the earthed electrode it appears to be brighter and is accompanied by a voluminous shower of discharge most conspicuous in fig. 1 (*b*). The main stroke generally diminishes in brightness as it recedes from the earthed electrode. In noting this increase of luminosity of both leader and main stroke near the earthed electrode it must not be forgotten that the film is more sensitive when slightly "fogged", and the voluminous shower of discharge present near the earthed electrode provides this fogging locally. There is, however, other evidence to show that the leader stroke sometimes truly increases in luminosity as it progresses: figs. 1 (*b*) and 1 (*c*) show a number of "stepped leaders" (Schonland, Malan and Collens 1935) which travel with high velocity until they catch up with the first leader, and thereafter the leader brilliancy is enhanced. Subsequent to the passage of the leader at any point the luminosity at that point falls to very low values in some records such as fig. 1 (*a*) near the high-voltage electrode, and fig. 1 (*b*) between the second and third leader strokes at the high-voltage electrode. In other records the luminosity does not fall to such an extent after the passage of the leader, as for instance in fig. 1 (*d*) (in this case the applied voltage is about twice the minimum spark-over voltage of the gap). (See Plate 6.)

The leader stroke is sometimes of a composite character as in the case of the stepped leaders to the lightning discharge. The first leader usually exhibits all the phenomena of branching and travels slowly: the subsequent leaders travel rapidly over the previously ionized main track and do not branch, but exhibit branching as they extend the previous track towards the opposite electrode (fig. 1 (*b*) is a good example of this). Where they blaze a new portion of track their velocity is comparable with that of the first leader. The number of stepped leaders seems to be very variable even under the same conditions of sparkover; the total time between first leader and main stroke is also very variable. For example, of two discharges across a 135 cm. gap one has a first leader-main stroke separation of 94 μsec. and exhibits no stepped leaders, the other has a separation of 180 μsec. and exhibits four stepped leaders. Generally the number of stepped leaders increases with total time to spark-over and a very rough average time between leader stroke components is 15 μsec., during which time the part of the stepped leader advancing over a fresh part of the gap traverses about 15 cm. For the purpose of determining the velocity of the leader stroke, measurements have been made of the time interval between the leader stroke and the main stroke at many points between the electrodes. The results of a typical analysis for a positive discharge are shown on the left of fig. 2 with the time interval as abscissa. The curved line represents the tip of the leader stroke; from this line the velocity at any point may be deduced, and is shown on the right of fig. 2. This discharge was accompanied by five stepped leader strokes which are shown in fig. 2 as lines parallel to the ordinate. The factors governing velocity of propagation will be discussed later.

Very infrequently the positive leader does not bridge the gap completely, and presumably a negative leader is propagated from the earthed plane to meet the downward-moving positive leader, although the time interval between such a leader and the main stroke cannot be observed and so must be less than 0·5 μsec. This type of discharge is shown in fig. 1 (*e*) where the junction point is approximately 20 % of the electrode spacing above the earthed plane. No discharge has yet been observed in which the height of the junction point above the earthed plane exceeds 30 % of the electrode spacing. The development of this discharge is graphically represented in fig. 3, and is exceptionally interesting in that the stepped leaders are seen to diminish in speed after they have reached the previous leader. Thus the leader "envelope" does not progress at a steadily increasing speed, but rather in bounds, each of which is initiated by the fresh pulse of energy from the successive stepped leaders. It is possible that the leader to the discharge which has been analysed in fig. 2 progresses in a similar manner, but the changes in velocity may be so small as to be undetectable.

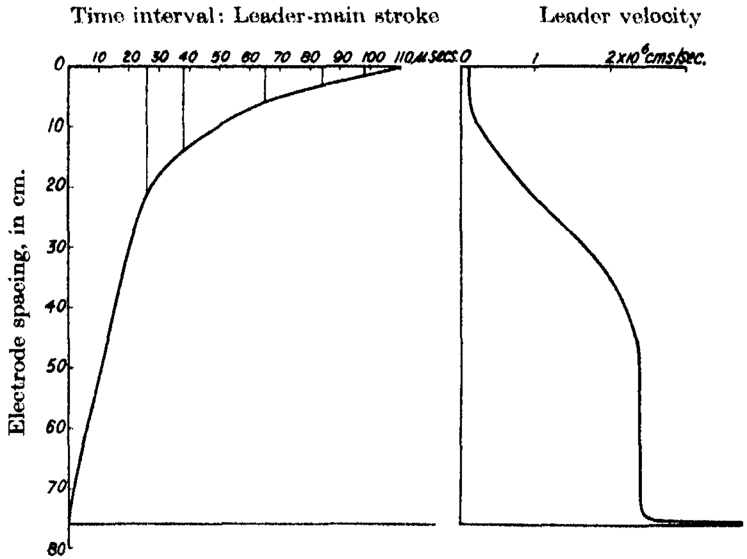


FIG. 2. Positive-point electrode: earthed plane electrode.
Development of a typical discharge.

The main stroke of the spark discharge like that of the lightning flash does not branch so frequently as the leader stroke, but only follows the more prominent of the leader stroke branches. This is clearly shown in fig. 1 (*d*) for two of the main branches. The intensities of the main stroke along branches are lower than those of the leader branch; they are also weak compared with the main strokes which show no diminution of intensity as they pass a junction point. At these points, however, the main stroke shows very localized "beads" of light, a few of which can be seen in fig. 1, and often these correspond to similar "beads" on the forks of the leader stroke (see fig. 1 (*d*)). Some of the main stroke branches exhibit a "comet" type of discharge where they subdivide; such a case is given in fig. 4, a drawing from an enlargement of a very short section of the spark: similar "comets" are visible in figs. 1 (*a*) and 1 (*e*). The speed of development of the main stroke and of its branches must be of the same order of magnitude. As this speed is too high to be measured no proof is available as to the direction of propagation of the main stroke, but the presumption that it travels upwards is based on the lightning analogue and on the fact that the intensely ionized section of the leader stroke at the instant of commencement of the main stroke is at the earthed electrode, not at the high-voltage electrode.

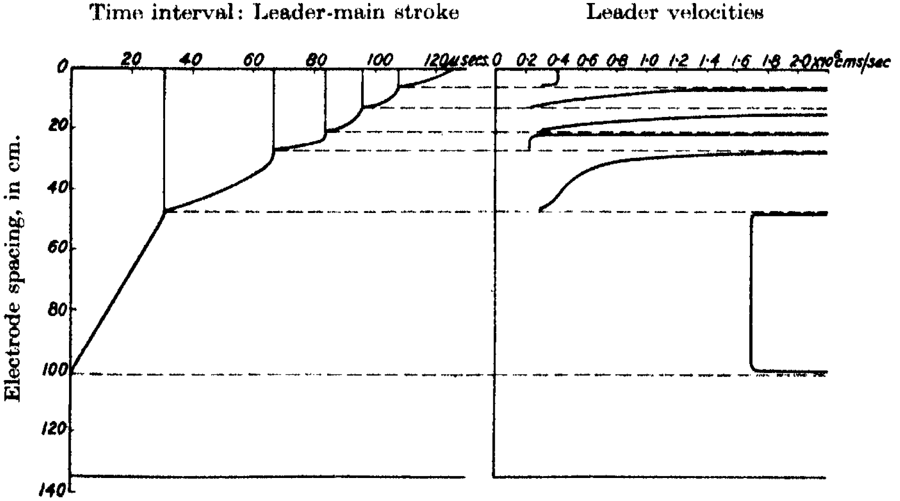


FIG. 3. Positive-point electrode: earthed plane electrode. Development of the stepped discharge shown in fig. 1 (*e*).

A number of oscillograms indicating the voltage at the point were taken in conjunction with the photographs; a typical oscillogram is shown in fig. 1 (*f*). Comparison of simultaneous records shows that the leader stroke is initiated at an early stage on the voltage wave front, while the subsequent wave shape deviates to a continually increasing extent from that on open circuit, due to the increasing voltage drop in the series resistance *R* as the leader stroke develops. Steps on the voltage wave may be observed, corresponding to the development of stepped leaders. There is a sudden drop in voltage immediately previous to spark-over, and this is found to occur simultaneously with the final stepped leader. In the case of the discharge corresponding to fig. 1 (*f*) this final leader occurred 6 μsec. before the main stroke.

(ii) *High-voltage positive-point electrode: pointed earthed electrode*

A typical record of this type of discharge is shown in fig. 5 (*a*) (Plate 7). The earthed point electrode projects out of an earthed plane for a distance equal to the interelectrode spacing. In addition to the downward directed leader stroke an upward directed leader stroke starts at the earthed point and the two meet, followed then by the main stroke propagated, presumably, in both directions. The upper leader and main stroke are similar to the discharges described in the previous section. The upward-directed leader exhibits upward branching which is later followed by upward branching of the main stroke, and where the leaders join there is an elongated section

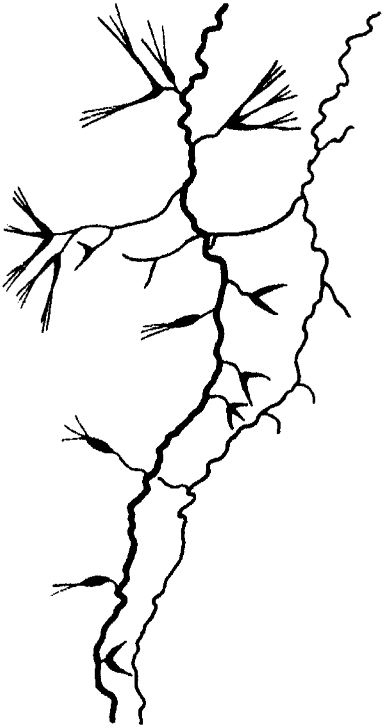
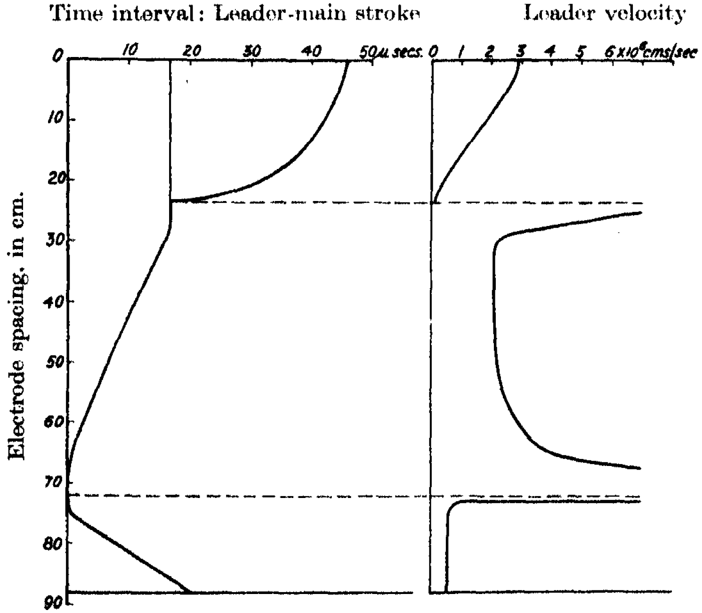


FIG. 4. Drawing to illustrate "comet" type discharge.

of intense discharge. The final leader-main stroke interval at the earthed point is 20 μsec. in fig. 5 (*a*) (Plate 7) and the interval at the high voltage point is 46 μsec.: thus the upward leader started when the downward leader had traversed about 30 % of the gap. An analysis of the leader developments of fig. 5 (*a*) is given in fig. 6. It will be seen that the velocity of the negative leader is less than one-third of that of the positive leader over the distances where the two leaders were developing simultaneously, and this is confirmed in the study of many records. The positive leader here has one or two steps and then progresses with uniform velocity. Between the parts of the leaders which are developing simultaneously the same voluminous shower of discharge (see fig. 5 (*a*)) takes place as is referred to in the foregoing section.


Time interval: Leader-main stroke

Leader velocity

FIG. 6. Positive-point electrode: earthed point electrode. Development of the stepped discharge shown in fig. 5 (*a*).

The presumption that the main stroke develops in both directions from the meeting point of the opposite leaders is again founded on the fact that the intensely ionized sections of the leaders at the moment of commencement of the main strokes is where they meet, not at either electrode: there is no visible discharge at either electrode for the 10 μsec. which precede the main stroke, so it is inconceivable that the main stroke starts from either electrode. A current-time oscillogram of this discharge is given in fig. 5 (*b*). This shows that the first leader at the earthed point takes place after 32 μsec. from the instant of application of the voltage, and a second (stepped) leader takes place 3 μsec. later, while the main stroke follows after a further 20 μsec. The agreement between the various leader-main stroke spacings and the current-time oscillograms is generally perfect, though not all the preliminary stepped leaders at the earthed electrode are strong enough to show on reproductions. It will be seen from the oscillograms that the leader currents rise and fall very rapidly, the maximum currents last for only 1 μsec. at most. Between the last leader and the main stroke some current is flowing from the ground electrode, about 30 % of the last leader stroke current, although the ionization produced by this current is not intense enough to be recorded photographically. The main-stroke current is determined by the series resistance and its duration by the circuit constants.

Branching of the discharge is enhanced by the presence of a number of earthed points. This is illustrated in fig. 5 (*c*), when two points projected 10 cm. out of the earthed plane (the two points are equidistant from the high-voltage point and lie in the same vertical plane). The positive leader divides after travelling about one-quarter of the gap; one of the branches is met by an upward-growing negative leader from one of the earthed points, and forms the ionized path over which the main stroke occurs. The other branch, which develops in the direction of the second point, terminates in mid-gap. Its path is retraced by the main stroke, and measurement of the time interval at different positions in the gap indicates that the speed of this branch leader is the same as that of the main leader after each has progressed equal distances. The development of the positive-branch leader is accompanied by an upward-growing negative leader from the second point.

(iii) *High-voltage positive spherical electrode: pointed earthed electrode*

The upper electrode had a diameter of 25 cm. and the electrode spacing was 75 cm. The results are generally similar to those in the foregoing section except that the negative upward-developing leader was lengthened by the partial suppression of the positive leader. The two leaders joined about 30 % of the gap spacing above the earthed electrode.

(iv) *High-voltage positive spherical electrode: plane earthed electrode*

The upper electrode had a diameter of 12·5 cm. and the electrode spacing was 75 cm. Whereas there is seldom an upward-developing leader from a plane electrode when the anode is a point, substitution by a 12·5 cm. sphere results apparently in a short, upward, negative leader, the presence of which can be inferred, since an intense junction point of the kind described in § 4 (ii) occurs at about 10 % of the gap spacing above the plane. This leader cannot, however, be distinguished on the photographs: obviously its time separation from the main stroke cannot exceed 0·5 μsec. and its speed of development must be high.

(v) *High-voltage positive spherical electrode: spherical earthed electrode*

Both spheres had a diameter of 25 cm.: electrode spacings of 25 and 75 cm. were employed. At the larger spacing the positive leader almost bridged the gap, the junction-point was only 5–10 % above the earthed sphere. The positive leader strokes are not accompanied by stepped leaders and the velocity of propagation is more or less uniform at $2 \cdot 5 \times 10^6$ cm./sec. From the upper spherical electrode straight shafts of discharge frequently extend, branching only after proceeding some 5–10 cm. and the leader proceeds from one of these branches. At the smaller spacing these shafts are longer and nearly bridge the gap, although the leader which initiates spark-over seldom develops from the end of one of these straight shafts. Figs. 7 (*a*) and 7 (*b*) (Plate 7) illustrate the nature of the discharge at 75 and 25 cm. spacing respectively (the band across the centre of fig. 7 (*a*) is a photographic fault).

5. THE NEGATIVE DISCHARGE

(i) *High-voltage negative-point electrode: plane earthed electrode*

Typical records of this type of discharge are shown in fig. 8 (Plate 8). The distinctive feature is that in addition to the negative leader which starts from the cathode, a positive leader rises out of the plane and meets the negative leader, an inversion of the phenomena described in § 4 (ii). If discharges from the edge of the plane are to be avoided the gap must be less than 70 % of the minimum width of the plane: the illustrations given in fig. 8 are for sparks definitely terminating on the smooth flat portion of the sheet-metal plane. Identical results have also been obtained by substituting a large shallow trough of water 3 ft. square for the sheet-steel plane: the upward-directed positive leader always rises from the plane to meet the downward-moving negative leader, and the two meet at varying positions usually rather above the middle of the gap. Both leaders branch in the directions of their propagation, and the same voluminous discharge surrounds the leaders in the vicinity of their meeting point as noted for the positive discharge between two points.

The discharge shown in fig. 8 (*a*) is across a 70 cm. gap, and in fig. 8 (*b*) across a 37 cm. gap. It will be seen that there are a number of brush discharges rising from the plane in addition to the leader to the main stroke. There is also a considerable amount of discharge at the high-voltage point electrode in the form of sharply defined streamers prior to the development of the leader stroke which meets the positive leader stroke: the leader stroke follows the ionized path blazed by the pre-discharges, but it differs greatly in appearance from the stepped leaders of the positive discharge as the pre-discharges are not connected to it, and the envelope of the tips of the pre-discharges is not continuous as in fig. 1 (Plate 6). These discharges will therefore be referred to as "pre-discharges" after which only one "leader" stroke occurs to initiate the main stroke.

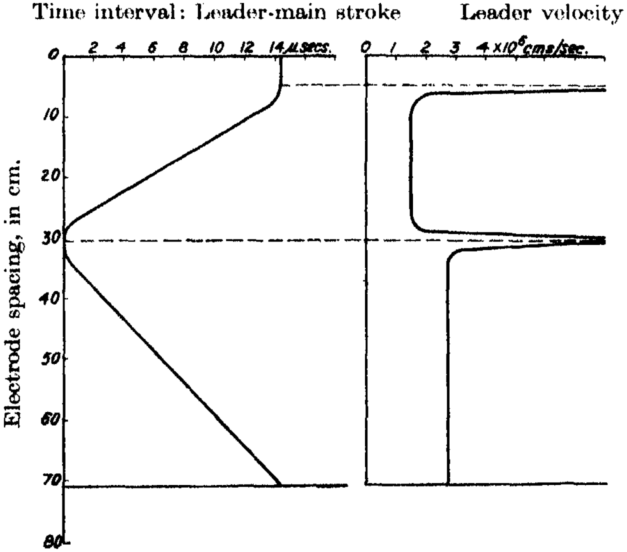


Fig. 9. Negative-point electrode: earthed plane electrode. Development of the discharge shown in fig. 8 (*a*).

An analysis of the development of the discharge shown in fig. 8 (*a*) is given in fig. 9, and indicates a negative leader velocity of $1 \cdot 5 \times 10^6$ cm./sec., and a positive leader velocity of $2 \cdot 6 \times 10^6$ cm./sec. The velocity of the negative leader is very high over the region previously ionized by a "pre-discharge", thereafter it is smaller than the positive-leader velocity. During this latter portion of the leader development the leader is not so sharply defined and its luminosity is maintained until the main stroke occurs—for a time of 10–15 µsec. at least—but during the initial portion of the leader stroke where the leader is traversing a pre-ionized path the luminosity only lasts about 1 μsec., and no visible discharge is apparent for the 10–15 μsec. separating the leader from the main stroke.

Fig. 8 (*a*) (Plate 8) exhibits a very interesting phenomenon at the junction point of the two leaders. There is a sphere of discharge 7 mm. in diameter, the luminosity of which lasts for 700 μsec. This phenomenon is not typical of the junction points of leaders and no explanation can be advanced for its infrequent appearance. The junction point in fig. 8 (*b*) is of the more usual form: it is generally elongated to a few cm. and has no specially long luminosity. Many kinks or bends of leader strokes show luminosity for longer times than the remainder of the leader. The electrodes remain incandescent for times far longer than the duration of the luminous discharge between the electrodes. This latter hardly ever exceeds 500 μsec., whereas the former is frequently 2000 μsec. The ground electrode shows the longest luminosity which might be explained by the fact that it is of iron, whereas the upper electrode is of brass, and the iron spectrum is notably strong in the ultra-violet.

The straight shaft of discharge from the positive electrode, previously observed for sphere-sphere breakdown (§ 4 (v)), is also present in a few photographs of the discharge from a negative point to earthed plane. Such a photograph is shown in fig. 8 (*c*), where the shaft is seen to develop from the earthed positive plane. The shaft discharge is very intense for about 15 cm. of the gap, and from this intense portion heavy branching develops. The first branches which occur suddenly at about 4 cm. above the plane are curved upwards in well-defined "parabolic" paths. The main stroke follows the path of the shaft discharge for about 6 cm., i.e. to above the first branching point; up to this point the speed of the shaft discharge is too fast to measure, and although the luminosity of the shaft is initially strong yet it falls rapidly, so that the space on the photograph between the shaft and the main stroke is quite clear. After this point, however, a normal positive leader stroke develops from the shaft discharge, at a speed of $3{\cdot}2 \times 10^6$ cm./sec., and its luminosity is maintained until the main stroke occurs. No pre-discharges are visible prior to the development of the shaft discharge, so that the latter cannot be considered as similar to a stepped leader such as described in § 4 (i). It appears therefore that the positive leader, of the shaft discharge type, can develop at a speed too high to be measured by the present camera, although it is traversing a path which has not been previously ionized. This fast positive leader has only been observed to develop from planar or spherical electrodes and not from point electrodes.

Oscillograms of two negative-impulse voltages are given in fig. 8 (*d*), one an open circuit record, the other showing spark-over. Simultaneous records were taken with the rotating camera, and it was found that the small steps on the voltage oscillograms coincide in time with the pre-discharges at the high-voltage electrode. In the event of spark-over the last step on the oscillogram occurs at the same instant as the final negative leader, and is followed by a large drop in voltage before the main spark-over takes place. If there is only corona discharge the same small steps corresponding to pre-discharges are visible, while there is a definite step-up in voltage when the pre-discharges cease.

(ii) *High-voltage negative-point electrode: pointed earthed electrode*

Typical records of this type of discharge are shown in fig. 10 (Plate 9): the spacing of the electrodes is 75 cm., the earthed point projects out of a plane for 2 % of the spacing in figs. 10 (*a*) and 10 (*b*), and for 100 % of the spacing for 10 (*c*) (Plate 9). The records show the two leaders meeting in the mid-gap region, on the average a little nearer to the high-voltage electrode than for a point-plane discharge. The individual pre-discharges at the high-voltage point are very conspicuous in figs. 10 (*a*) and 10 (*b*) (only the part of the discharge down to the junction of the leaders is shown in fig. 10 (*b*)) and a notable characteristic of the negative discharge is the sharp bifurcation visible in the pre-discharges. The intensity of the two forks is frequently greater than that before bifurcation, and in this resembles the negative Lichtenberg figures. Another characteristic of the negative discharge is also conspicuous in figs. 10 (*a*) and in 10 (*c*), referred to as an "explosive discharge": from the high-voltage electrode in 10 (*c*), and from halfway down the leader in 10 (*a*), a straight shaft of discharge develops for some 10 cm. terminated by an intense elongated discharge 3–8 cm. in length from which a shower of clearly defined negative streamers proceed towards the earthed electrode. A strange feature of this discharge is that its track is not followed by the subsequent main stroke. Immediately after the explosive discharge has left the negative leader track the velocity of this track diminishes rapidly. The explosive discharge is *never* met by the ascending positive leader. A graphical representation of fig. 10 (*a*) is given in fig. 11.

It has been said that the main stroke often traverses the more important of the branches to the leader stroke. A notable example of this is seen at the cathode of fig. 10 (*b*). The first of the pre-discharges is composed of two separate tracks, one to right the other to left at an angle of about 120° to the first. The second pre-discharge takes the path of the left-hand track, the third, fourth and fifth pre-discharges and the main leader take the path of the right-hand track, but when the main discharge reaches the cathode it also traverses the left-hand track: an interval of 94 μsec. separates the second pre-discharge from the main stroke.

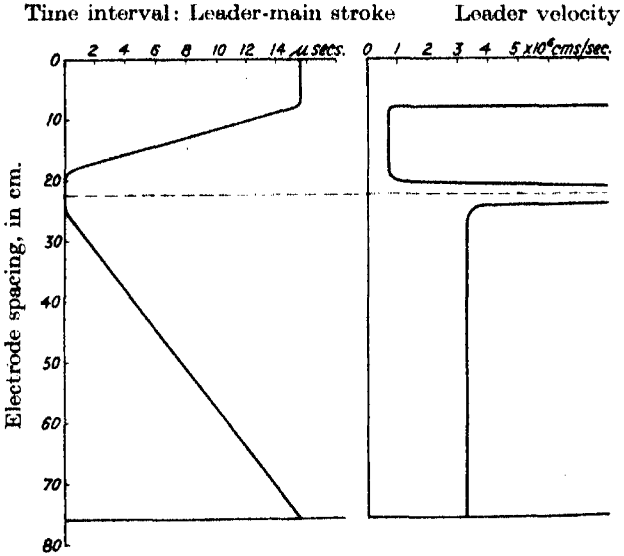


FIG. 11. Negative-point electrode: earthed point electrode. Development of the discharge shown in fig. 10 (*a*).

Fig. 10 (*c*) is accompanied by the current oscillogram 10 (*d*) taken at the earthed electrode. The time on the oscillogram from application of the impulse to commencement of the upward leader is 102 μsec. and 23 μsec. elapse between leader and main stroke; the leader-main stroke separation on the rotating film is 23 μsec. which is in agreement with the oscillographic records.

(iii) *High-voltage negative sphere electrode: pointed earthed electrode*

Two diameters of spherical electrode were employed, 25 and 100 cm.; the point/25 cm. sphere spacing was 45 cm., and the point/100 cm. sphere spacing was 100 cm.

The negative leader is very short from the 25 cm. sphere and is almost non-existent from the 100 cm. sphere where the positive leader (illustrated in fig. 12 (*a*), Plate 10) bridges at least 95 % of the whole gap from the earthed point to the sphere. The current oscillogram for this is given in fig. 12 (*b*). The positive leader is stepped, and its structure at the anode shown photographically corresponds to the details of the current oscillogram: this is further illustrated in the drawing of fig. 13 which is accompanied

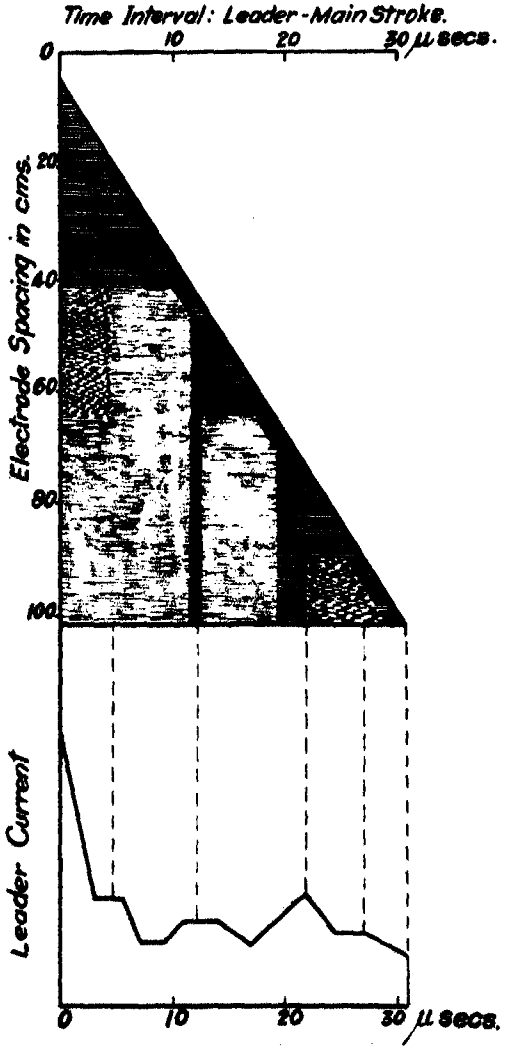


Fig. 13. Negative-sphere electrode: earthed point electrode. Development of the discharge and the synchronized current oscillogram shown in fig. 12.

by an analysis of the current oscillogram. The velocity of the positive leader is $3 \cdot 1 \times 10^6$ cm./sec. and is uniform over the whole gap, in marked distinction to the velocities of the downward-positive leaders in figs. 2, 3 and 6, although two stepped leaders join the initial leader. The stepped leaders correspond

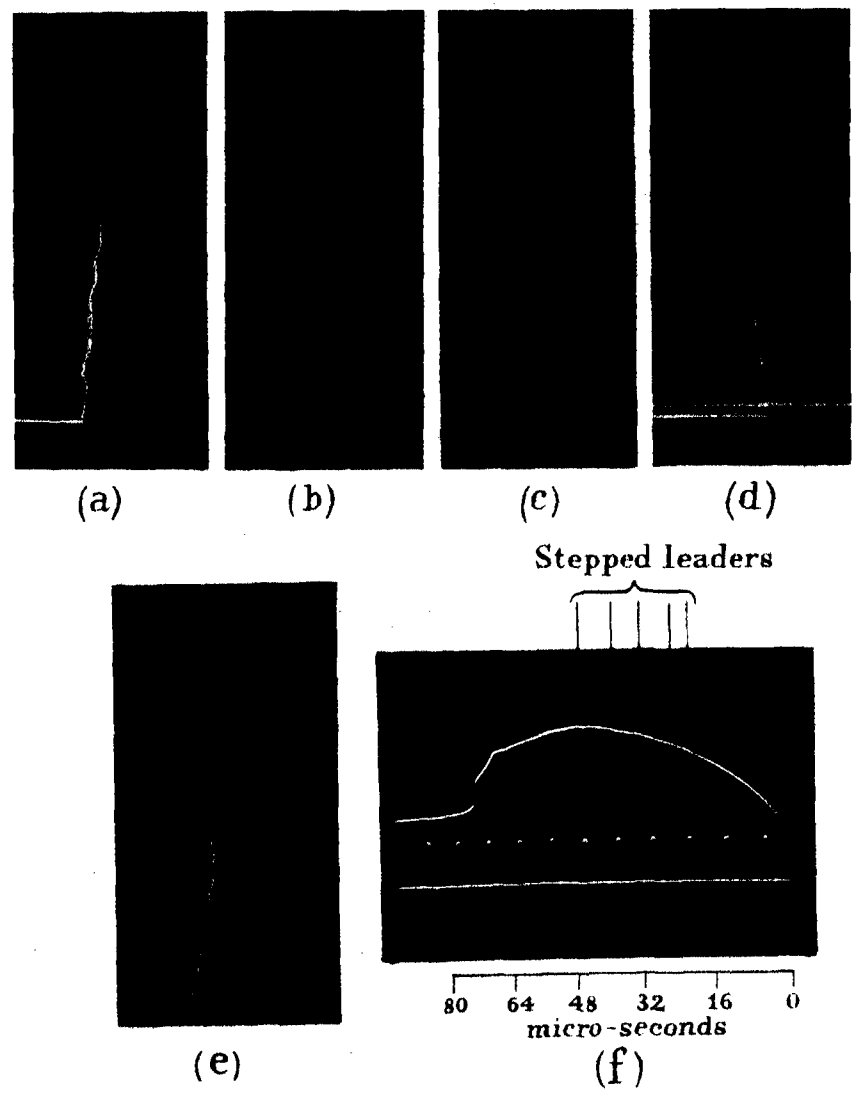


FIG. 1. Positive-point electrode: earthed plane electrode.

	Gap length	Series resistance
(*a*)	d = 183 cm.;	R = 1,040,000 ohms.
(*b*)	d = 135 cm.;	R = 1,040,000 ohms.
(*c*)	d = 76 cm.;	R = 1,040,000 ohms.
(*d*)	d = 76 cm.;	R = 1,040,000 ohms.
(*e*)	d = 135 cm.;	R = 1,040,000 ohms.
(*f*)	d = 76 cm.;	R = 500,000 ohms.

(*Facing p.* 112)

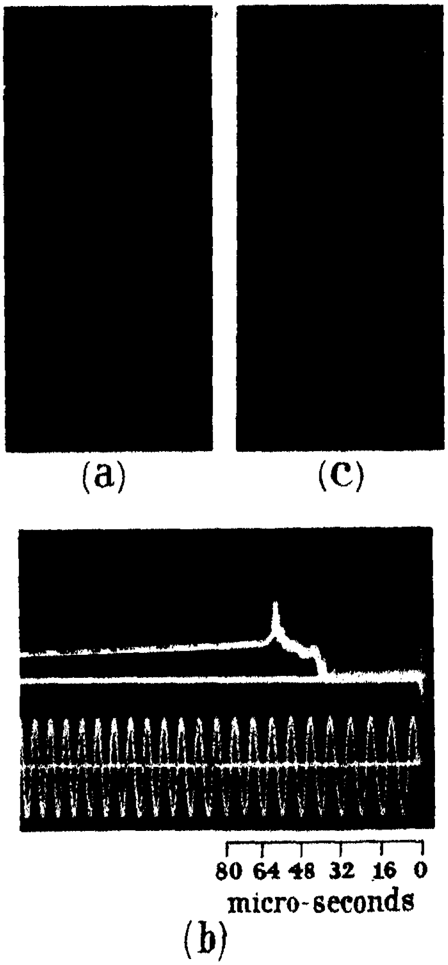


Fig. 5

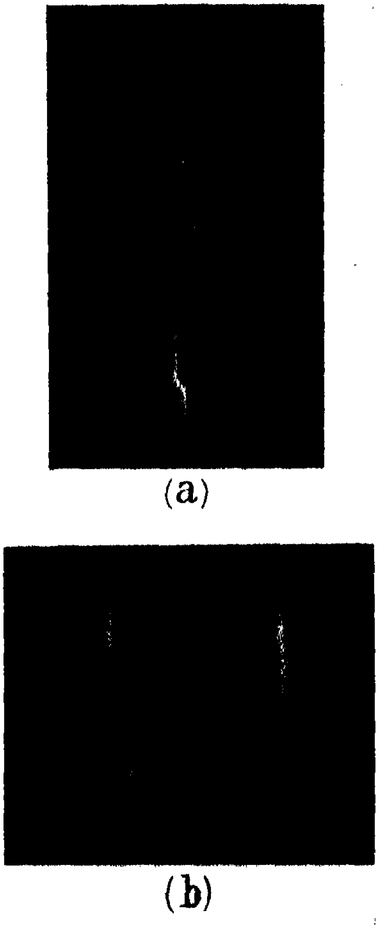
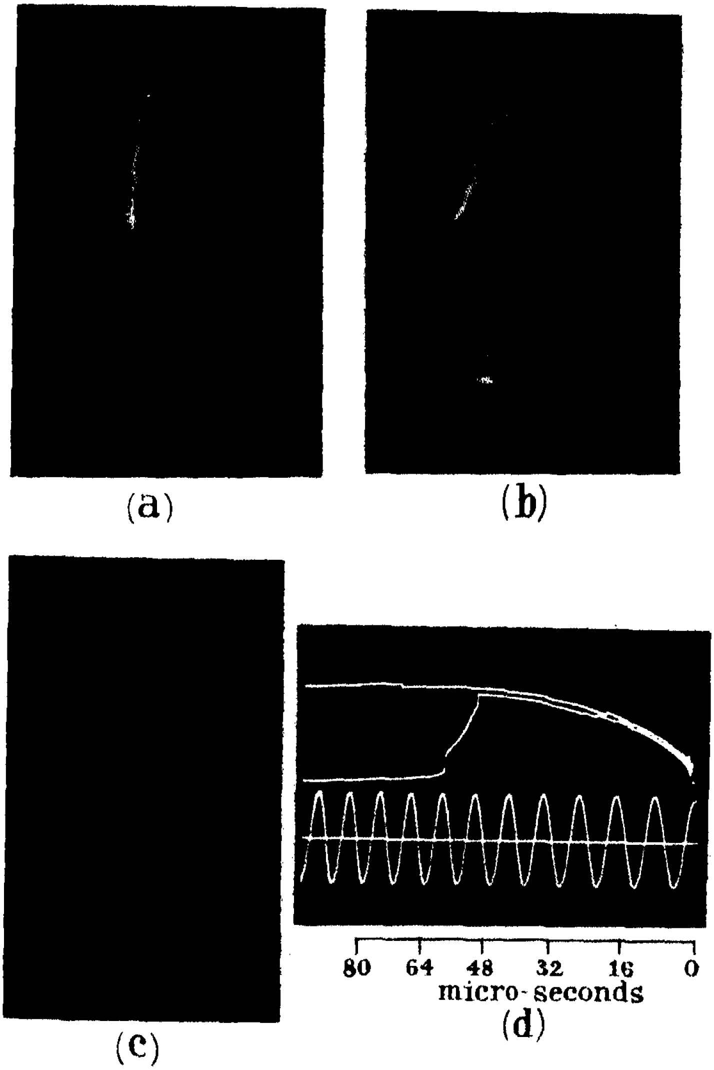


Fig. 7

Fig. 5. Positive-point electrode: earthed point electrode. (*a*) $d = 89$ cm.; $p = 68$ cm.; $R = 480{,}000$ ohms. (*b*) Current oscillogram synchronized with (*a*). (*c*) $d = 94$ cm.; $p = 10$ cm.; $R = 400{,}000$ ohms.

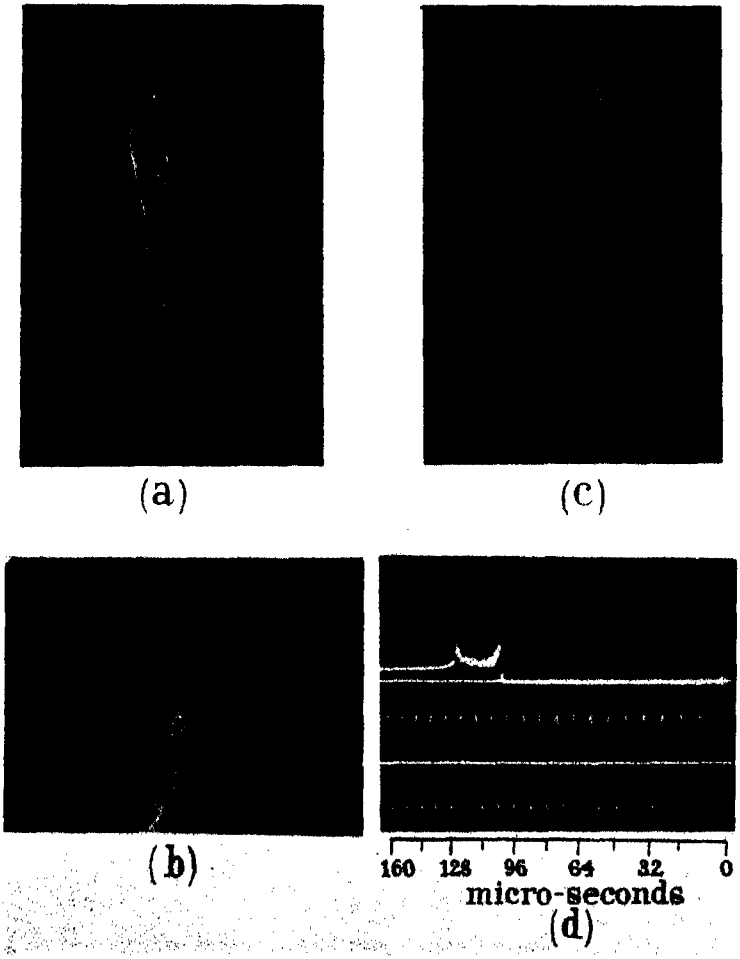
Fig. 7. Positive-sphere electrode: earthed sphere electrode. (*a*) $d = 75$ cm.; $R = 300{,}000$ ohms. (*b*) $d = 25$ cm.; $R = 300{,}000$ ohms.



(a) (b)

(c) (d)

Fig. 8. Negative-point electrode: earthed plane electrode. (*a*) $d = 70$ cm.; $R = 400{,}000$ ohms. (*b*) $d = 70$ cm.; $R = 400{,}000$ ohms. (*c*) $d = 38$ cm.; $R = 1{,}900{,}000$ ohms. (*d*) $d = 51$ cm.; $R = 500{,}000$ ohms. (Voltage oscillogram.)



(a) (c)

(b) (d)

Fig. 10. Negative-point electrode: earthed point electrode. (*a*) $d = 76$ cm.; $p = 1 \cdot 25$ cm.; $R = 1,900,000$ ohms.; (*b*) $d = 76$ cm.; $p = 1 \cdot 25$ cm.; $R = 1,900,000$ ohms. (*c*) $d = 76$ cm.; $p = 76$ cm.; $R = 1,300,000$ ohms. (*d*) Current oscillogram synchronized with (*c*).

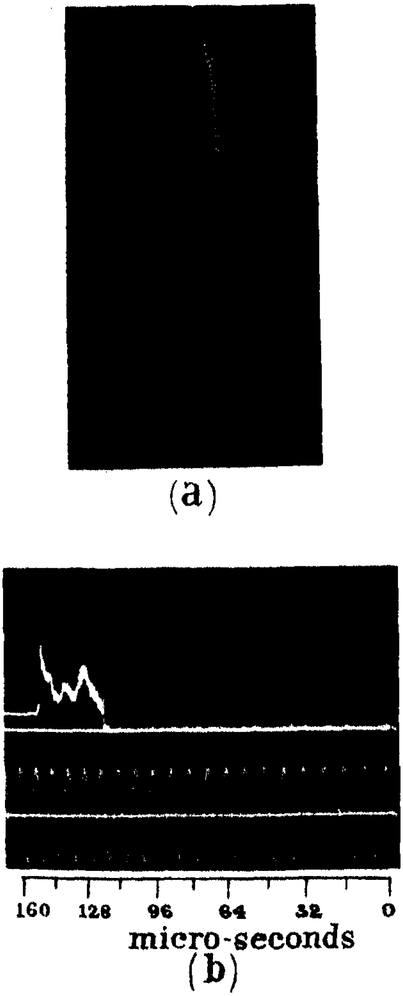


Fig. 12

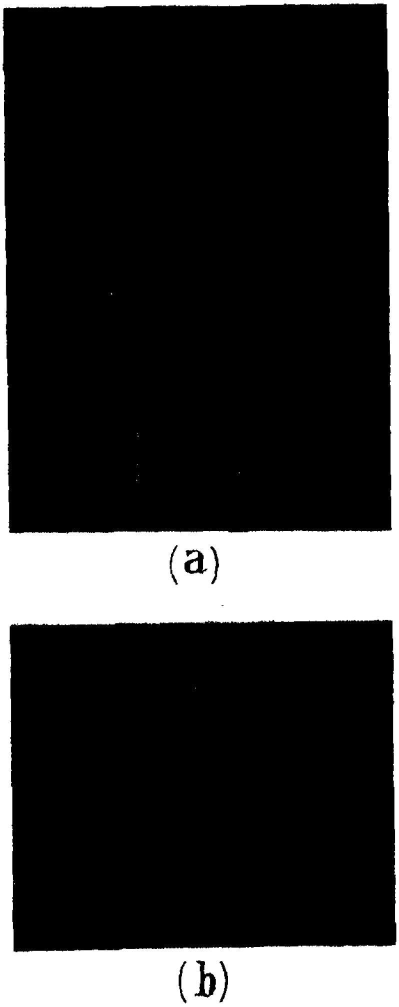


Fig. 14

Fig. 12. Negative-sphere electrode: earthed point electrode. (*a*) $d = 100$ cm.; $p = 50$ cm.; $R = 1,040,000$ ohms. (*b*) Current oscillogram synchronized with (*a*).

Fig. 14. Negative-sphere electrode: earthed sphere electrode. (*a*) $d = 75$ cm.; $R = 300,000$ ohms. (*b*) $d = 25$ cm.; $R = 300,000$ ohms.

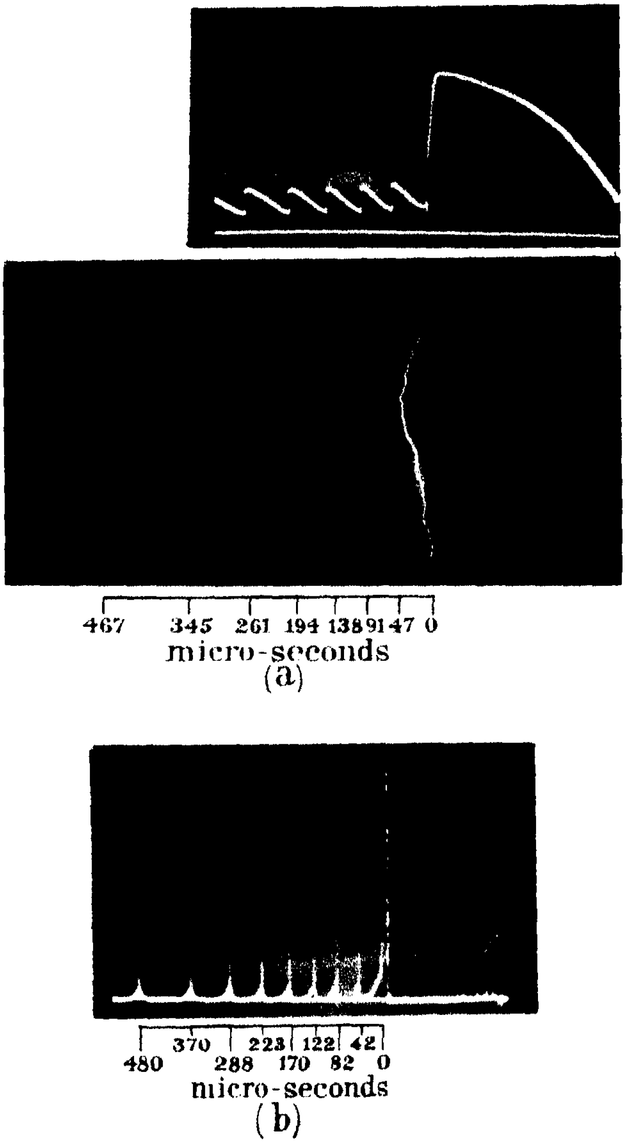


FIG. 18. Multiple-stroke discharge for a 63 cm. gap between a positive-point electrode and an earthed point electrode ($p = 66$ cm.). (*a*) Photograph of discharge together with synchronized voltage oscillogram. (*b*) Current oscillogram for a negative discharge over the same gap.

to the peaks on the current oscillograms, and have velocities too high to be measured; the duration of luminosity of these steps is from 1 to 3 μsec. Where they join the initial leader the duration of luminosity of the leader is considerably increased up to 7 μsec. indicating an increase of ionization at the moment of junction of the leaders. As the upper electrode is approached a large shower of discharge develops similar to that formed at the earthed electrode in fig. 1 (*b*). The direction of development cannot be stated with certainty: it probably develops simultaneously from the sphere—on which short darts can be seen—and from the many branches of the upward growing leader.

(iv) *High-voltage negative spherical electrode: plane earthed electrode*

The results are similar to those reported under § 5 (i), the only sphere diameter used was 12·5 cm. at a spacing of 75 cm. from the plane, so that the electrode can almost be considered as a point electrode. The junction point of the opposing leader strokes is about in the centre of the gap.

(v) *High-voltage negative spherical electrode: spherical earthed electrode*

The spheres of 25 cm. diameter were spaced at 25 and 75 cm. At the large spacings the opposing leader strokes meet at about the centre of the gap; at the diameter spacing the meeting point varies from the centre up to the high-voltage sphere. Breakdown for the two spacings is illustrated in figs. 14 (*a*) and 14 (*b*) (Plate 10). The leaders usually start from the spherical electrodes with very high velocity, then branch, and develop further with more or less uniform velocities of the order of $1{\cdot}5 \times 10^6$ cm./sec. for the negative leader and 4×10^6 cm./sec. for the positive leader. The straight shaft of discharge from the earthed sphere in fig. 14 (*a*) is comparable with that in fig. 7 (Plate 7) from the positive sphere.

6. FACTORS INFLUENCING SPARK DEVELOPMENT

(*a*) *Circuit resistance and gap spacing*

In the work previously reported (Allibone and Schonland 1934) an impulse generator having a discharge capacitance of 1000 μμF was employed. To make the discharge circuit aperiodic a 1000 ohm (approx.) resistance was inserted. The resultant main discharge photographed through a glass lens was of such an intensity that the leader stroke 0·2–1 mm. ahead of it could readily be distinguished, but the composite character of the leader stroke was not appreciated. The velocities of positive and negative leaders calculated on the assumption that each traversed the whole gap from high-voltage electrode to ground with uniform velocity were 9 and 15×10^6 cm./sec. respectively. In the present work a generator having a discharge capacitance of 10,000 $\mu\mu$F was employed together with the quartz lens, and with only 1000 ohms series resistance the record of the main stroke was so intense that it obscured the leader stroke. In consequence only those records taken with circuit resistances from 9000 ohms to the upper limit of 1·9 megohms have proved suitable for analysis.

(i) *Positive high-voltage electrode*. The velocities of positive leader strokes can be measured with considerable accuracy as the leaders are nearly always continuous. The values are remarkably consistent for the same gap and same circuit conditions. In general, as shown in fig. 2, the velocity increases with increasing distance from the anode, and this is true for all values of circuit resistance. The change of velocities of stepped leaders, as shown in figs. 3 and 6, is also a phenomenon independent of circuit resistance. For a gap spacing of 76 cm. (point/plane electrode) the time between the initiation of the leader and the occurrence of the main stroke increases with increase of resistance as shown in fig. 15. Isolated points have also been plotted for gaps of 38, 135 and 183 cm., and indicate an inappreciable variation in time intervals over this range of gap lengths for high values of series resistance. However, oscillographic measurements show that the time to breakdown increases with gap length at low values of series resistance, and for 1000 ohms the values given in fig. 15 are taken from such measurements (Allibone, Hawley and Perry 1934; Bellaschi and Teague 1934).

In making comparison between leader-stroke velocities an arbitrary distinction has been drawn between the initial velocity and that more or less uniform velocity which the leader acquires as it approaches the opposite electrode; this value is described as the "final" velocity, and for convenience it is measured at a point ¾ of the gap spacing from the high-voltage electrode. The variation of these velocities with resistance is given in fig. 16 for a gap length of 76 cm., and a few isolated values for other gap lengths are also given.

The initial velocity of the leader decreases with increase in series resistance, and velocities as low as $0{\cdot}2 \times 10^6$ cm./sec. have been measured for a resistance of 1·45 megohms. Although it is probable that even lower initial velocities can be obtained if the resistance is further increased, yet these velocities become more difficult to measure due to the increased number of short-stepped leaders. Not only is the intensity of these stepped

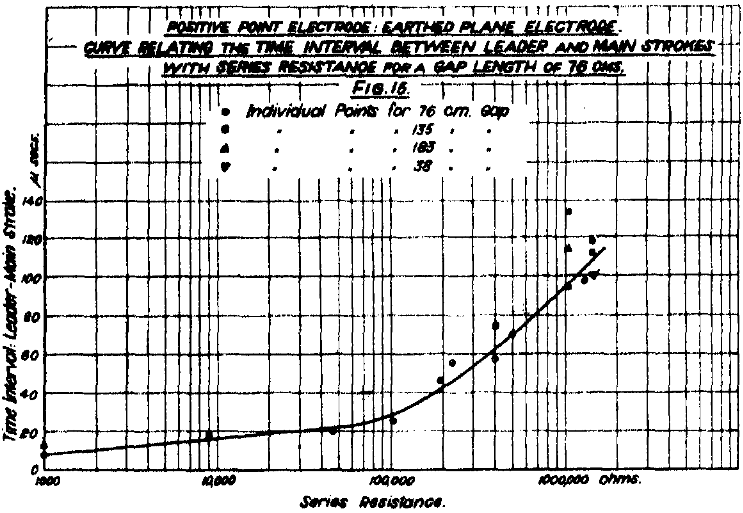


Fig. 15. Positive-point electrode: earthed plane electrode. Curve relating the time interval between leader and main strokes with series resistance for a gap length of 76 cm.

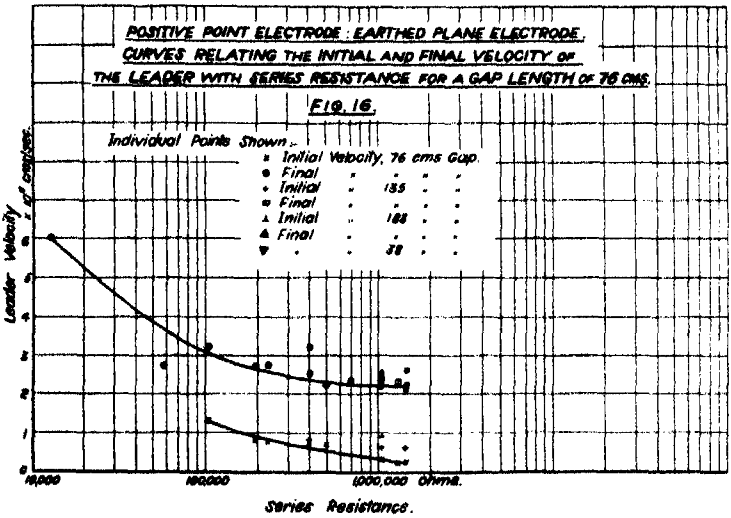


Fig. 16. Positive-point electrode: earthed plane electrode. Curves relating the initial and final velocities of the leader with series resistance for a gap length of 76 cm.

leaders very weak when the resistance is in excess of 1·5 megohms, but the envelope of their tips is not clearly defined. It is probable that this reduction in intensity of the leader stroke is responsible for the decreased speed of the leader when the series resistance is increased, but in addition the increase in series resistance causes a decrease in the rate of rise of voltage at the high-voltage electrode, and this will be a further factor tending to increase the time to breakdown and presumably to decrease the initial leader velocity.

The final velocity appears to be more or less independent of gap length at high values of series resistance (e.g. 1·45 megohms): it varies from $2{\cdot}6 \times 10^6$ cm./sec. for a 183 cm. gap to $2{\cdot}2 \times 10^6$ cm./sec. for a 76 cm. gap and $2{\cdot}1 \times 10^6$ cm./sec. for a 38 cm. gap. The decided flattening of the final velocity curves as the resistance is increased suggests that there is a lower limit to the velocity at which the leader will progress without apparent reinforcement from stepped leaders. Below 10,000 ohms it is not possible to measure the final velocity with the present camera due not only to the fact that the time interval between the leader and the main stroke is reduced, but also to the confusion caused by the increased luminosity of the main stroke.

The development of the positive leader does not appear to be affected in any way when the earthed plane is replaced by a point electrode. An insufficient number of records have been taken to enable curves to be drawn, but it will be seen from the analysis of such a discharge given in fig. 6 that the final velocity of the positive leader is $2{\cdot}3 \times 10^6$ cm./sec., which is in accord with the point-plane records. This record was made for a series resistance of 480,000 ohms, and the analysis of two other discharges (gap = 68 cm.: length of earthed point = 22 cm.; series resistance = 1,300,000 ohms) gave a final leader velocity of $2{\cdot}2 \times 10^6$ cm./sec.

The velocity of the upward-growing negative leader is affected considerably by change in resistance. In the record analysed in fig. 6 its velocity is $0{\cdot}6 \times 10^6$, whereas for the other examples quoted above, where the resistance is 1,300,000 ohms, it is only $0{\cdot}1 \times 10^6$ cm./sec.

(ii) *Negative high-voltage electrode.* The mechanism of the discharge when the high-voltage electrode is negative is not consistent, even under identical circuit conditions. Sometimes there are fast pre-discharges followed by a fast final leader from which an explosive discharge may or may not occur; at other times there are no pre-discharges and the explosive discharge develops directly out of the high-voltage electrode itself. The velocity of the negative leader after the explosive discharge has occurred can be measured, but its value is variable over a considerable range for the same group of records. Thus for a 57 cm. gap between a negative high-voltage point and an earthed point projecting 1·25 cm. out of an earthed plane ($R = 1·9$ megohms), the velocities of the negative leader ranged from $0·8 \times 10^6$ to $0·3 \times 10^6$ cm./sec. Velocities varying from $3·2 \times 10^6$ to $1·1 \times 10^6$ cm./sec. were obtained for a 92 cm. gap with a series resistance of 0·4 megohm (the earthed point projecting 11 cm. out of the earthed plane). There is seen to be an increase in the negative leader velocity when the resistance is decreased, but as yet an insufficient number of records have been taken for the same gap to give curves for the variation of leader velocity with resistance. The time between the *final* negative leader and the main stroke varied between 11 and 20 μsec. for the 57 cm. gap ($R = 1·9$ megohms) and between 9 and 16 μsec. for the 92 cm. gap ($R = 0·4$ megohm). Thus the time interval does not vary appreciably over this wide range of series resistance, but it must decrease considerably when the resistance is reduced to 1000 ohms as the time-to-spark-over is only 5–6 μsec. for the above-mentioned gaps (Allibone, Hawley and Perry 1934).

The upward-growing positive leader is continuous in its development, and, in the majority of the records examined, is initiated almost at the same instant as the final negative leader. In no case does it precede the final negative leader. The velocity of the positive leader is found to decrease with increase in resistance, though as in the case of the negative leader the velocity may vary by a factor of 3 under identical circuit conditions: for high values of series resistance the mean velocity is found to be in close agreement with the final velocity of the downward-growing leader from a positive point; for low resistances the upward positive leader develops faster than the downward positive leader.

(*b*) *Magnitude of applied voltage*

In all the foregoing records (with the exception of fig. 1 (*d*)) the applied voltage was just sufficient to cause spark-over. The effect of over-voltages is to decrease the time-to-spark-over, decrease the number of pre-discharges, increase the luminosity of the leader strokes and the duration of their luminosity, increase the amount of branching and increase the velocity of propagation of the leader strokes. Fig. 17 shows the change of formation and of velocity of the positive leader for a 76 cm. gap as the voltage is raised from the minimum spark-over value to twice this value. It will be noticed that the greatest change of velocity with voltage occurs at the beginning of the leader, the final velocity only increases by 25 % when the applied voltage is doubled.

(c) *Nature of applied voltage*

In all the foregoing records the voltage has been applied rapidly to the gap, the rate of rise of voltage being of the order of 10–1000 kV/μsec. With such rapid rates of rise of voltage no appreciable space charge can be created before the leader stroke mechanism develops. To test what effect such space charge may have on leader stroke formation some experiments were performed with a source of constant potential of negative polarity (300 kV) the voltage being slowly increased until spark-over occurred; a series resistance of 80,000–500,000 ohms was in circuit. The electrode arrangement chosen was the point/point gap, and from the points intense brush discharges were maintained for at least 10 sec. before spark-over occurred. The general appearance of the D.C. spark discharge was found to be similar to that of the impulse spark discharge of the same polarity (such as fig. 10 (*c*), Plate 9) except that the D.C. leader stroke showed no pre-discharges at the high-voltage electrode, and the upward-growing positive leader appeared to develop from a point 5–10 cm. above the earthed point, apparently from the limit of the visible corona at that point. The velocity of propagation

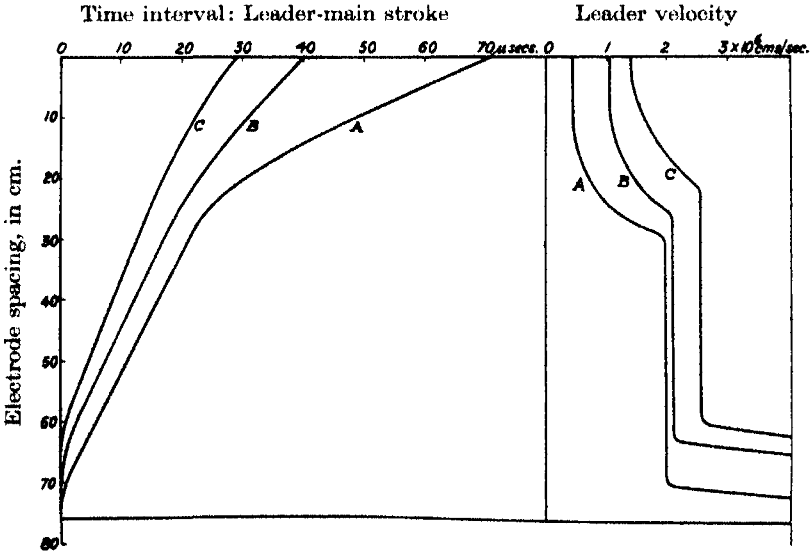


FIG. 17. Effect of over-voltage on leader development for a positive point to earthed-plane gap. $R = 500{,}000$ ohms.

A. Minimum spark-over voltage ($= V$).

B. Voltage $= 1{\cdot}25$ V.

C. Voltage $= 2{\cdot}0$ V.

of the downward negative and upward positive leaders were also comparable with those of the impulse discharge (negative, $0{\cdot}2$–$0{\cdot}3 \times 10^6$ cm./sec.: positive, 1–3×10^6 cm./sec.). A further experiment was performed with a group of 20 points projecting from an earthed plane giving an extensive layer of space charge over the plane. The results did not differ materially from the foregoing results so that it can be said that the position of the junction points of the leader strokes and their velocities are not affected by the presence of space charge or the field distortion produced thereby.

(*d*) *Nature and resistance of the earthed electrode*

Since lightning discharges may terminate on earth of widely differing resistivity, photographs were taken of impulse voltage and constant voltage spark discharges between a high-voltage point electrode and a plane covered with 5 in. of dry sand and (*a*) earthed direct, or (*b*) earthed through a resistance of 30,000 ohms. No difference could be detected between the results so obtained and those relating to earthed metallic electrodes: the nature of the ground electrode does not therefore influence the speed of leader stroke development.

7. MULTIPLE STROKES

The lightning flash can consist of a succession of separate strokes: as many as 27 have been observed (Schonland, Malan and Collens 1935). This feature may be reproduced in the spark discharge by a very simple method. A condenser of 36 $\mu\mu$F capacitance was connected in parallel with the discharge gap, through a resistance of 814 ohms, sufficient to make the circuit aperiodic. A high resistance ($R = 1{\cdot}8$ megohms) was connected between the impulse generator and the high-voltage electrode of the discharge gap. The resultant discharge, shown in fig. 18 (*a*), consists of a series of strokes, the first of which is preceded by a leader stroke. A synchronized oscillogram illustrating the variation of voltage at the high-voltage electrode is also given in fig. 18 (*a*) and a current oscillogram relating to another similar discharge is shown in fig. 18 (*b*); unfortunately it was not possible to obtain simultaneous voltage and current oscillograms as only one oscillograph was available. (See Plate 11.)

It will be observed from the photograph of the discharge that the branching of the leader is followed by branching of the first main stroke. The subsequent strokes do not branch, are less intense than the first stroke and follow its main path in detail. Similar effects have been noted in the lightning flash, where it has been found that the first stroke is usually more intense than subsequent strokes, and is always much more heavily branched, while the strokes all follow the same main path. The luminosity of the first stroke is not apparent after 5 μsec. in the discharge of fig. 18 (*a*), but it is evident that the degree of ionization must still be considerable when the next stroke occurs, 47 μsec. after the first, since the track followed by this stroke is identical with that of the first stroke. There is no indication of leaders to strokes subsequent to the first, though it may be that these would be revealed by a camera of higher speed.

The intensities of the successive strokes are all less than that of the first stroke, and vary among themselves according to the time intervals between them. As in the lightning discharge a stroke which succeeds a previous stroke after a long time interval is more intense than a stroke which has occurred after a short time interval. This is to be expected, since the shorter the time interval the less will be the voltage developed across the condenser in parallel with the gap, and hence the lower the energy available when a stroke occurs. In the case of one particular discharge the second stroke occurred 37 μsec. after the first main stroke and required a voltage of $0{\cdot}21\,V$, where V is the voltage necessary to cause initial breakdown. The third stroke occurred after a further 41 μsec, at a voltage of $0{\cdot}24\,V$ and was correspondingly more intense than the second stroke.

The development of the leader to the first stroke of the discharge shown in fig. 18 (*a*) is affected considerably by the presence of the condenser in parallel with the gap. There is not the pronounced stepping which was previously evident for high values of series resistance, and the luminosity persists until the main stroke occurs. The speed of the downward growing positive leader is *uniform* at $3{\cdot}0 \times 10^6$ cm./sec., and that of the upward growing negative leader is uniform at $1{\cdot}5 \times 10^6$ c./sec.; both these speeds show an increase on those which would have been obtained if the parallel capacitance had not been present. When the polarity of the high-voltage electrode is negative the downward-growing negative leader meets the upward-growing positive leader at about one-third the gap length beneath the high-voltage cathode. The leaders remain luminous until the main stroke occurs; the negative leader velocity decreases from $3{\cdot}2 \times 10^6$ cm./sec. as it leaves the cathode, to $1{\cdot}6 \times 10^6$ cm./sec. as it approaches the junction point of the two leaders. The ascending positive leader develops at a uniform rate of $4{\cdot}1 \times 10^6$ cm./sec. There are no differences between the positive and negative discharges as regards multiple strokes, other than that of leader formation, and hence the details given above for the positive multiple discharge are applicable to both polarities.

8. DISCUSSION OF RESULTS

(a) *Structure of the spark*

The most interesting feature of the present work is the proved existence of the positive as well as the negative leader stroke to the main discharge. Hitherto the leader stroke to the lightning flash has been presumed (not always proved) to be associated with a negative discharge to earth and a mechanism similar to that suggested by Dorsey (1926) has been advanced (Schonland and others 1935) to explain the propagation of the negative leader.

A further interesting feature is the continuous nature of the leader stroke. The leader to the first stroke of a lightning flash is always stepped, and Schonland (1937) could only offer strong presumptive evidence for the existence of a "pilot" streamer which travels in front of the stepped streamers preparing an ionized path for them. The spark photographs prove the existence of the continuous pilot streamer, reinforced from time to time by stepped leaders which travel rapidly from the electrode to the pilot streamer: the stepped leaders to the lightning flash cannot always be traced from the pilot streamer back to the cloud.

A remarkable characteristic of the negative discharge is that it is always accompanied by a positive discharge which develops from the earthed electrode whether of point, plane or spherical form, and the two leaders meet in the mid-gap region. This is true also for power frequency A.C. sparks and for D.C. sparks. If this is a general characteristic of all negative discharges in the laboratory it might be presumed to characterize the lightning flash also: some few photographs of flashes taken with stationary cameras show the thickening of the flash, reminiscent of the junction point of the spark leaders (see Jensen 1933, right-hand flash of fig. 21 positive field-change reported, i.e. negative cloud base), and a few photographs show upward branching of the lightning flash indicating upward propagation of a leader (see Jensen 1934). In this connexion it might be noted, as previously stated (Allibone and Meek 1937) that all spark photographs show that the direction of branching is the direction of propagation of the leader stroke.

(b) *Velocity of propagation of the leader*

It is of interest to compare the speed of propagation of the leader stroke with the speed of spark formation as calculated from time-lag observations by other workers. In the following table calculated speeds and, where possible, actual speeds are quoted for gaps varying from 1 mm. to 3 km.

It will be seen that the mean speed of propagation varies by only a factor of × 10 for gaps from 0·1 to 150 cm. for minimum impulse spark-over, and is almost no higher for the lightning flash. The voltage of the lightning flash is not necessarily the minimum voltage which would result in spark-over from cloud to ground, and it has been shown that overvoltages can cause an increase in leader velocity. The positive leaders are generally continuous and their velocities increase as the leaders approach the cathode; the total range of velocities is from 0·2 to 20×10^6 cm./sec. The negative leaders are composed of very fast and very slow sections, the very fast part adjacent to the cathode travels too fast to be measured and is generally traversing a preionized track, the very slow section develops from the tip of the fast leader at speeds of 0·1 to 4×10^6 cm./sec. until it meets an ascending positive leader. Time-lag studies naturally only give the average values quoted in the table.

(c) *The mechanism of the leader stroke*

The oscillograms of the current in the spark clearly show that the leader stroke starting from the electrode carries a current of the order of 1 amp. for less than 1 μsec., after which the current at the electrode diminishes until the main stroke or stepped leader stroke occurs. The intensity of the leader stroke does not vary greatly as it traverses the gap so that it may be presumed to contain the same density of ion pairs over the whole of its path. The leader appears therefore as a column of ionized air the front of which is advancing due to fresh ions being formed at its tip, while the rear of the column recedes from the electrode due to recombination of the ions. The propagation of the positive discharge has been discussed by Simpson (1926): his suggestions require modification to include recombination of ions. The modification suggested is illustrated in fig. 19 (*a*). Ionization occurs at the electrode tip due to the concentration of field there. Electrons occurring by chance or liberated photoelectrically ahead of the streamer travel towards it ionizing as they move in the intense field. If almost no movement of the positive ions occurs the tip of the ionized region will advance towards the cathode although the motion of the ionizing particles—electrons—is towards the anode. The velocity of the electrons inside the channel is governed by the voltage gradient in the channel and may be expected to be small compared with the velocity of the ionizing electrons at the tip of the channel, and compared with the velocity of propagation of the channel. The channel does not therefore become filled with positive ions as pictured by Simpson but remains a region of ionized gas of almost equal number of positive ions and electrons which combine according to

TABLE I. SPEEDS OF SPARK FORMATION

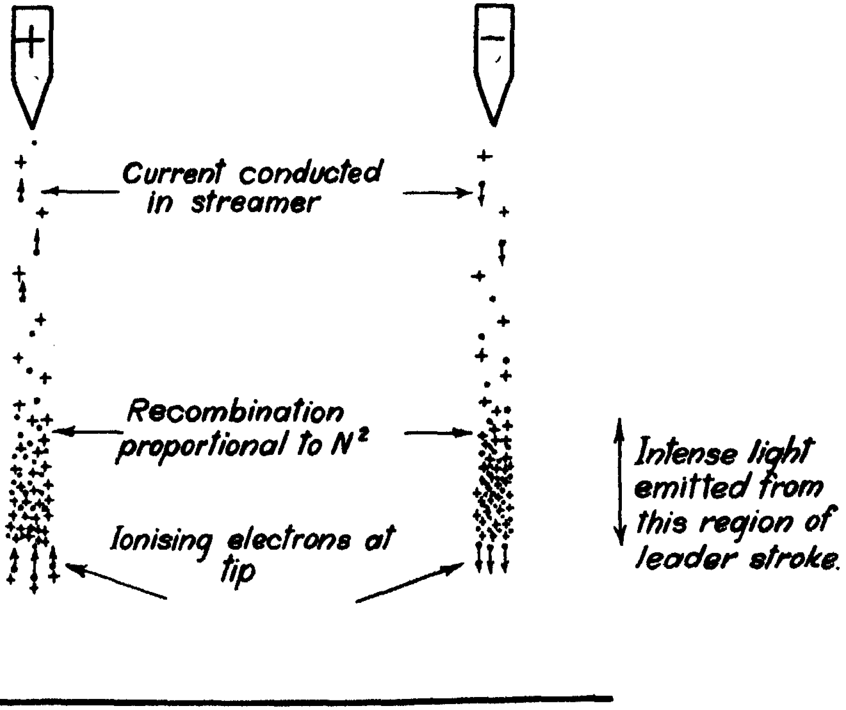
(The speeds quoted have been obtained by dividing the gap length by the time to spark-over, except in those cases where direct measurement is possible)

Gap length cm.	Time lag $\times 10^{-8}$ sec.	Average speed $\times 10^{8}$ cm./sec.	Over-voltage	Polarity of high-voltage electrode	Type of electrode	Observer	
0·1	400	0·25	Min. S.O.	Not given	Parallel plates	Rogowski 1928	
0·1	50	2·0	40%	Not given	Parallel plates	Rogowski 1928	
0·1	15–25	6–4	Not given	Not given	Brass spheres, 2 cm. dia.	Beams 1928	
0·2	25–50	8–4	Not given	Not given	Brass spheres, 2 cm. dia.	Beams 1928	
0·3	140	2	10%	−ve	—	White 1936	
0·3	40	7·5	40%	−ve	—	White 1936	
0·3	70	4	10%	−ve	Brass spheres, ¾ in. dia.	Newman 1937	
0·3	5	60	200%	−ve	Brass spheres, ¾ in. dia.	Newman 1937	
0·5	14	3·5	½%	Not given	Brass spheres, 4 cm. dia.	Dunnington 1931	
1·5	300	5	Min. S.O.	+ve & −ve	Point/point (30° cones)	Burawoy 1926	
25·0	6	4	Min. S.O.	+ve	H.V. point/earthed plane	Bellaschi and Teague 1934	
25·0	3·5	7	10%	+ve	H.V. point/earthed plane	Bellaschi and Teague 1934	
25·0	4	6*	Min. S.O.	−ve	H.V. point/earthed plane	Bellaschi and Teague 1934	
25·0	2	12*	10%	−ve	H.V. point/earthed plane	Bellaschi and Teague 1934	
38·0	—	2·1	Min. S.O.	+ve	H.V. point/earthed plane	$R=1,450,000$	Allibone and Meek 1938
46·0	—	1·5 (−ve leader) 4·0 (+ve leader)	Min. S.O.	−ve	H.V. point/earthed plane	$R=450,000$	
71·0	—	1·5 (−ve leader) 2·6 (+ve leader)	Min. S.O.	−ve	H.V. point/earthed plane	$R=400,000$	
76·0	—	2·2	Min. S.O.	+ve	H.V. point/earthed plane	$R=1,450,000$	
105	—	2·6	Min. S.O.	+ve	H.V. point/earthed plane	$R=1,450,000$	
135	—	2·6	Min. S.O.	+ve	H.V. point/earthed plane	$R=1,450,000$	
183	—	2·5	Min. S.O.	+ve	H.V. point/earthed plane	$R=1,040,000$	
150	12	12	Min. S.O.	+ve	H.V. point/earthed plane	Bellaschi and Teague 1934	
150	6	25	10%	+ve	H.V. point/earthed plane	Bellaschi and Teague 1934	
150	9	17*	Min. S.O.	−ve	H.V. point/earthed plane	Bellaschi and Teague 1934	
150	3	50*	10%	−ve	H.V. point/earthed plane	Bellaschi and Teague 1934	
100,000–300,000	—	10–30	Not known	−ve	Cloud/ground	Schonland 1935	

* These speeds are certainly too high as no allowance is made for the positive leader which develops from the earthed plane.

White (1934) measured cathode streamer velocities of $1{\cdot}4 \times 10^7$ cm./sec. in nitrogen for an 8 mm. gap: no statement is made regarding minimum spark-over voltage.

the usual law $dn/dt = -\alpha n^2$ (where $\alpha = 1{\cdot}5 \times 10^{-6}$) unless ionization is maintained by current passing through the channel. From a knowledge of the current in the leader—1 amp.—and the diameter—1 mm.—the density of ion pairs must be of the order of 10^{15} ions/c.c. and this number would be reduced by $\sqrt{10}$—and the luminosity by 10—in 10^{-9} sec. unless replenished. It is not possible to state for how long the luminosity persists:


+
−
Current conducted in streamer
Recombination proportional to N^2
Ionising electrons at tip
Intense light emitted from this region of leader stroke.

(a) (b)

FIG. 19. Drawing to illustrate the suggested development of leader strokes.

an upper limit of $0{\cdot}5 \times 10^{-6}$ sec. applies to some of the leader strokes. In others—and especially with overvoltages—the luminosity, and the current recorded in the leader does not fall so that ionization proceeds within the channel due to the voltage gradient along it.

When the value of the series resistance R is large the leader stroke current flowing to the anode causes a reduction in its potential, and the leader stroke travels with diminishing velocity; after a few microseconds the potential of the anode rises again and a second leader develops rapidly along the weakly ionized path of the first leader until it joins the first leader and rapidly augments its velocity. As shown in the earlier sections several of these stepped leaders may leave the anode, and the current oscillograms completely confirm this idea.

The propagation of the long negative discharge has been discussed by Dorsey (1926), Schonland and Collens (1934) and Cravath and Loeb (1935). The latter two based their discussions on the observed lightning "leader" stroke, average velocity of $8 \cdot 10^8$ cm./sec. but it is realized now that this figure related in 1934 to a "dart" leader, not to the leader to the first stroke, the velocity of which was only measured in 1935 (Schonland and others 1935) and found to average $1 \cdot 5 \times 10^7$ cm./sec. This velocity is comparable with the laboratory spark leader velocity whether of positive or negative polarity. The negative leader stroke shown in fig. 19 (*b*) advances as electrons in the vicinity of the tip move from the tip under the action of the high field. Although Cravath and Loeb have shown theoretically that the electron velocity and the stroke velocity can be very widely different, there is in fact no great difference between the leader stroke velocity and the velocity of an electron in the minimum ionizing field. Overvoltages produce faster leader strokes but also higher electron velocities. In the electric spark the advancing electrons must produce 10^{12} electrons/cm. path to give the observed leader stroke current: these would be produced by one electron travelling 1 cm. in a field of 34 kV/cm. ($\alpha = 28$ for $X/p = 50$; Posin 1936).

The only reason that can be advanced to explain why the negative leader develops more slowly than the positive leader is that behind the negative leader tip is a concentration of positive ions which partially neutralizes the field whereas a similar concentration of positive ions at the tip of the positive leader augments the field. No explanation can be offered for the explosive types of discharge which are under further investigation.

The authors wish to express their thanks to Dr W. Payman and the Safety in Mines Research Board for the loan of the Quartz Flint lens, to members of the staff of the High-Voltage Laboratory for assistance, and to Dr A. P. M. Fleming, C.B.E., Director and Manager of Research and Education Departments, Metropolitan-Vickers Electrical Co., Ltd., for permission to publish this paper.

9. SUMMARY

A general account is given of the development of the high-voltage electric spark, based on the study of over 1000 photographs taken with a rotating film camera.

The electric spark is shown to consist of two principal components, a leader stroke and a main stroke—analogous to the lightning flash. A leader stroke invariably starts from a positive electrode, and sometimes also starts from a negative electrode: sometimes the structure of the leader stroke is simple, sometimes it is of the "stepped" variety. The leader stroke is always branched at many places, and the direction of branching is the direction of its propagation in space: branching thus forms a criterion of the direction of leader-stroke development. The main stroke develops in the reverse direction to that of the leader stroke and at a velocity too high to be recorded. The velocities of the positive and negative leader strokes are of the order of 10^6–10^7 cm./sec., the positive leaders being the faster. Oscillograms of current and voltage support the photographic results. Multiple stroke discharges have been produced and show the characteristics of the multiple lightning flashes: the first is initiated by a slow leader stroke and is branched, the subsequent main strokes are either without leader strokes or the resolution of the camera is inadequate to show them: they are not branched.

REFERENCES

Allibone, T. E. 1938 *J. Instn Elect. Engrs.* In press.
Allibone, T. E., Hawley, W. G. and Perry, F. R. 1934 *J. Instn Elect. Engrs*, **75**, 670.
Allibone, T. E. and Meek, J. M. 1937 *Nature, Lond.*, **140**, 804.
Allibone, T. E. and Schonland, B. F. J. 1934 *Nature, Lond.*, **134**, 736.
Beams, J. W. 1928 *J. Franklin Inst.* **206**, 809.
Bellaschi, P. L. and Teague, W. L. 1934 *Trans. Amer. Inst. Elect. Engrs*, **53**, 1638–45.
Burawoy, O. 1926 *Arch. Elektrotech.* pp. 186–219.
Cravath, A. M. and Loeb, L. B. 1935 *Physics*, **6**, 125.
Dorsey, N. E. 1926 *J. Franklin Inst.* **201**, 485.
Dunnington, F. G. 1931 *Phys. Rev.* **38**, 1535.
Edwards, F. S. and Scoles, G. J. 1937 *Engineering*, **144**, 222.
Jensen, J. C. 1933 *J. Franklin Inst.* **216**, 707.
— 1934 *Sci. Mon., Lond.*, **39**, 560.
Newman, M. 1937 *Phys. Rev.* **52**, 652.
Posin, D. Q. 1936 *Phys. Rev.* **50**, 650.
Rogowski, W. 1928 *Arch. Elektrotech.* **20**, 186–219.
Schonland, B. F. J. 1937 *Trans. S. Afr. Inst. Elect. Engrs*, **28**, 204.
Schonland, B. F. J. and Collens, H. 1933 *Nature, Lond.*, **132**, 407.
— — 1934 *Proc. Roy. Soc.* A, **143**, 654.
Schonland, B. F. J., Malan, D. J. and Collens, H. 1935 *Proc. Roy. Soc.* A, **152**, 595.
Simpson, G. C. 1926 *Proc. Roy. Soc.* A, **111**, 56.
White, H. J. 1934 *Phys. Rev.* **46**, 99.
— 1936 *Phys. Rev.* **49**, 507.

Quantum theory of Einstein-Bose particles and nuclear interaction

By N. Kemmer

Beit Scientific Research Fellow. Imperial College of Science and Technology, London, S.W. 7

(*Communicated by S. Chapman, F.R.S.—Received* 9 *February* 1938)

1. Introduction

The description of nuclear interaction in terms of a neutron-proton "exchange force" appears to be well justified and generally accepted. However, there has hitherto been no satisfactory suggestion as to the nature of the field of charged particles, the virtual emission and reabsorption of which, according to Heisenberg's (1932) picture, would give rise to this type of interaction. It may be now considered certain that this field is not identical with the electron-neutrino field of Fermi's theory, the magnitude of nuclear forces being far too large to be compatible with the small empirical value of the constant of β-decay.

As an alternative and simpler description of the nuclear field Yukawa (1935) put forward the idea that the interaction is transmitted by charged particles obeying Einstein-Bose statistics. He showed that the resulting nuclear potential would be proportional to $r^{-1}\exp(-2\pi m_0 cr/h)$, m_0 being the rest mass of the Bose particles. Thus, forces of a correct range would be obtained with a rest mass about 100 times that of the electron.

The apparent discovery of particles ("heavy electrons") with a mass of this order of magnitude in cosmic radiation by Neddermeyer and Anderson (1937) has aroused considerable interest in Yukawa's suggestion, and various aspects of this possibility have been discussed by a number of authors (Yukawa 1935, 1937; Yukawa and Sakata 1937; Oppenheimer and Serber 1937; Stueckelberg 1937; Fröhlich and Heitler 1938; Kemmer 1938; Bhabha 1938*b*). In particular Yukawa and Sakata (1937) have studied one possible form of the theory in considerable detail.

It is the purpose of this paper to consider this theory from a more general point of view. There are various ways of generalizing Yukawa's treatment, the most important concerning the spin of the new particle. From the mechanism suggested it is clear that the spin must be taken to be integral, that is, $2n$ times that of the neutron or proton. Yukawa's equations are one way of describing the case $n = 0$. For this value there is a second alternative theory, and there are also two independent possibilities for $n = 1$. On the other hand, a consistent theory for higher values of n does not appear possible.

The expression for the interaction of the neutrons and protons with the Bose particles will also be taken in a very general form, a generalization being possible even in the case treated by Yukawa. It will be shown that the similarity of all the cases studied is so great that *a priori* it seems impossible to discriminate in favour of any of them. A decision can, however, be reached if one evaluates the neutron-proton force in all the different cases. Comparing with experimental data, only one possibility proves tenable, and it is of importance to note that this is a case with spin and not the scalar case hitherto usually considered.

In an accompanying paper by Fröhlich, Heitler and Kemmer (1938) this case has been selected for more detailed treatment. As far as comparison with experiment is possible in view of the present rather unsatisfactory state of quantum field theories, the account of known data given by the formalism proposed can be regarded as very good.

2. PROCA'S EQUATIONS AND THEIR QUANTIZATION

The wave equations for particles with a spin value twice that of the electron or proton have already been used in other investigations. In particular the theory proposed by Proca (1936) contains much of the formalism used in the following. As an introduction to the systematic development to follow in the next section, we will study and quantize the equations of Proca.† The interaction with heavy particles will not yet be considered here. For the scalar case used by Yukawa the analogous quantization had already been performed earlier by Pauli and Weisskopf (1934). Apart from the difference in the spin eigenvalue the results of the present procedure are found to be identical with theirs.

The wave equations of Proca can be written in the following form:

$$\frac{\partial \phi_\beta}{\partial x_\alpha} - \frac{\partial \phi_\alpha}{\partial x_\beta} = \kappa \chi_{\alpha\beta}, \tag{1}$$

$$\frac{\partial \chi^{\alpha\beta}}{\partial x_\alpha} = \kappa \phi^\beta. \tag{2}$$

---

† It has just come to the writer's notice that the question dealt with in this section has recently been treated independently by Durandin and Erschow (1937). Their results are obviously identical with ours, but it cannot be judged from the short publication available, whether their methods are the same. For methodical reasons we still prefer to include this section.

Here $x_0 = ct$, and for any tensor component $a_0 = -a^0$, $a_i = a^i$ $(i = 1, 2, 3)$. Greek suffixes are allowed all values from 0 to 3, whereas Latin suffixes run from 1 to 3 only. Unless they occur as summation variables the suffixes i, k and l will in future always denote the values 1, 2, 3 in cyclic permutation.

Proca's equations are obviously closely related to Maxwell's equations, the main difference being that the "potentials" ϕ_α recur on the right-hand side of the second set (2). Apart from that a difference consists in the fact that all quantities are here assumed to be complex.

We can immediately deduce the existence of a 4-vector that satisfies an *equation of continuity*, namely

$$s_\alpha = \frac{eci}{\hbar}(\chi^*_{\alpha\beta}\phi^\beta - \chi_{\alpha\beta}\phi^{\beta *}). \tag{3}$$

Thus s_α can be taken to represent the current density of our particles. Similarly, one can deduce the existence of an energy-momentum tensor. We will here only put down its 00-component, the energy density:

$$T_{00} = c\kappa(\phi_0^*\phi_0 + \phi_i^*\phi_i + \chi_{0i}^*\chi_{0i} + \tfrac{1}{2}\chi_{ik}^*\chi_{ik}). \tag{4}$$

It is satisfactory that this quantity is essentially positive.

The theory can now be easily stated in Hamiltonian form. Let us express T_{00} in terms of the six quantities ϕ_i and χ_{0i} together with their complex conjugates and their spatial derivatives alone. This can be done by using the equations

$$\kappa\chi_{ik} = \frac{\partial\phi_k}{\partial x_i} - \frac{\partial\phi_i}{\partial x_k} = (\mathrm{curl}\,\phi)_l \tag{5}$$

and

$$\kappa\phi_0 = -\frac{\partial\chi_{0i}}{\partial x_i} = -\mathrm{div}\,\chi_0 \tag{6}$$

and their complex conjugates. We thus obtain

$$H = \int T_{00}\,dV = c\kappa\int dV\left[|\vec{\phi}|^2 + |\vec{\chi}_0|^2 + \frac{1}{\kappa^2}|\mathrm{div}\,\chi_0|^2 + \frac{1}{\kappa^2}|\mathrm{curl}\,\phi|^2\right]. \tag{7}$$

Equations (5) and (6), which alone of equations (1) and (2) do not involve time derivatives, can be regarded as *definitions* of χ_{ik} and ϕ_0 respectively, and the latter quantities can in this way be completely eliminated as variables of our theory. The possibility of doing this for ϕ_0 is the essential difference from electrodynamics.

It is now seen immediately that we can obtain the remaining equations of (1) and (2) as the canonical equations of our Hamiltonian system. We must take the ϕ_i and ϕ_i^* to be the canonical co-ordinates q, the χ_{0i}^* and χ_{0i} being the corresponding momenta p.

The equation $\partial H/\partial p = \dot{q}$ here takes the form

$$\frac{\partial \phi_i}{\partial x_0} = \kappa \chi_{0i} - \frac{1}{\kappa}\frac{\partial}{\partial x_i}\operatorname{div}\chi_0, \tag{8}$$

and from the relation $\partial H/\partial q = -\dot{p}$ we obtain

$$-\frac{\partial \chi_{0i}}{\partial x_0} = \kappa \phi_i + \frac{1}{\kappa}\frac{\partial}{\partial x_k}\operatorname{curl}_l \phi - \frac{1}{\kappa}\frac{\partial}{\partial x_l}\operatorname{curl}_k \phi. \tag{9}$$

With the use of the definitions (5) and (6) we thus get

$$\frac{\partial \phi_i}{\partial x_0} - \frac{\partial \phi_0}{\partial x_i} = \kappa \chi_{0i} \tag{10}$$

and

$$\frac{\partial \chi^{\alpha k}}{\partial x_\alpha} = \kappa \phi^k. \tag{11}$$

Similarly, the conjugate complex relations are derived. Including (5) and (6) we have the complete set of Proca's equations.

Having found the Hamiltonian form, quantization is simple. We postulate the commutation rules

$$[\phi_i(x), \chi_{0k}^*(x')] = [\phi_i^*(x), \chi_{0k}(x')] = -\frac{\hbar}{i}\delta_{ik}\delta(x-x'), \tag{12}$$

while all other pairs of quantities are taken to commute. Their invariance can be proved by the usual methods. Assuming (12), the equations of motion (8) and (9) can be derived by means of the relations

$$[H, \phi_i] = \frac{\hbar}{i}\dot{\phi}_i \quad \text{and} \quad [H, \chi_{0i}] = \frac{\hbar}{i}\dot{\chi}_{0i}. \tag{13}$$

and their conjugates.

It should be observed that by making use of (6) we can here avoid the special considerations for ϕ_0 that are necessary in electrodynamics. This represents a considerable simplification.†

We will now divide each field variable into a "transversal" and a "longitudinal" part:

$$\phi_i = \psi_i + \Pi_i^* \tag{14}$$

with

$$\operatorname{div}\psi = 0 \quad \text{and} \quad \operatorname{curl}\Pi^* = 0,$$

---

† In proving the invariance of (12) it is however initially necessary to proceed as in radiation theory, considering ϕ_0 as a separate variable. This form of the theory has been dealt with in more detail by Bhabha (1938*b*).

and similarly
$$\chi^*_{0i} = \pi_i - \Psi^*_i \tag{15}$$

with
$$\operatorname{div}\pi = 0 \quad \text{and} \quad \operatorname{curl}\Psi^* = 0.$$

The Hamiltonian can now be represented by

$$H = c\kappa\int dV\left[\,|\vec{\psi}\,|^2+|\vec{\pi}\,|^2+\frac{1}{\kappa^2}|\operatorname{curl}\psi\,|^2+|\vec{\Psi}\,|^2+|\vec{\Pi}\,|^2+\frac{1}{\kappa^2}|\operatorname{div}\Psi\,|^2\right], \tag{16}$$

with the commutation rules

$$\begin{aligned}&[\pi_i(x),\psi_k(x')] = [\pi^*_i(x),\psi^*_k(x')]\\ &\qquad = [\Pi_i(x),\Psi_k(x')] = [\Pi^*_i(x),\Psi^*_k(x')] = \frac{\hbar}{i}\delta_{ik}\delta(x-x'),\end{aligned} \tag{17}$$

all other combinations commuting. We transform to momentum space, putting

$$\left.\begin{aligned}&\psi_i = \frac{1}{\sqrt{L^3}}\sum_k q_i(k)\exp(-i\vec{k}\,\vec{x}\,), \quad \pi_i = \frac{1}{\sqrt{L^3}}\sum_k p_i(k)\exp(+i\vec{k}\,\vec{x}\,),\\ &\qquad\qquad\qquad\qquad\qquad\qquad\left(k_i = \frac{2\pi n_i}{L},\ n_i = \text{integer}\right)\\ &\Psi_i = \frac{1}{\sqrt{L^3}}\sum_k Q_i(k)\exp(-i\vec{k}\,\vec{x}\,), \quad \Pi_i = \frac{1}{\sqrt{L^3}}\sum_k P_i(k)\exp(+i\vec{k}\,\vec{x}\,),\end{aligned}\right\} \tag{18}$$

and denoting the complex conjugates correspondingly. If we now choose the directions of the co-ordinate axes for each Fourier component separately, we can do this in such a way that

$$q_3 = Q_1 = Q_2 = p_3 = P_1 = P_2 = 0.$$

Further we put

$$Q_3 = q^{(0)}, \quad \frac{1}{\sqrt{2}}(q_1+iq_2) = q^{(+1)}, \quad \frac{1}{\sqrt{2}}(q_1-iq_2) = q^{(-1)},$$

$$P_3 = p^{(0)}, \quad \frac{1}{\sqrt{2}}(p_1-ip_2) = p^{(+1)}, \quad \frac{1}{\sqrt{2}}(p_1+ip_2) = p^{(-1)}.$$

The energy then becomes

$$H = c\kappa\sum_k\sum_{\epsilon=-1}^{+1}\left[q^{(\epsilon)*}(k)\,q^{(\epsilon)}(k)\left(1+\frac{k^2}{\kappa^2}\right)+p^{(\epsilon)*}(k)\,p^{(\epsilon)}(k)\right]. \tag{19}$$

Similarly, the total charge can be shown to be

$$\frac{1}{c}\int s^0\,dV = -\frac{ei}{\hbar}\sum_k\sum_{\epsilon=-1}^{+1}[p^{(\epsilon)}(k)\,q^{(\epsilon)}(k) - p^{(\epsilon)*}(k)\,q^{(\epsilon)*}(k)]. \tag{20}$$

We perform yet another transformation of the variables, following the procedure of Pauli and Weisskopf (1934) in the scalar case. We put

$$\left.\begin{aligned} p^{(\epsilon)}(k) &= \left(\frac{\hbar}{2\kappa}\right)^{\frac{1}{2}}(k^2+\kappa^2)^{\frac{1}{4}}\,[a_\epsilon^*(k)+b_\epsilon(k)],\\ q^{(\epsilon)}(k) &= \left(\frac{\hbar\kappa}{2}\right)^{\frac{1}{2}}\frac{i}{(k^2+\kappa^2)^{\frac{1}{4}}}\,[a_\epsilon(k)-b_\epsilon^*(k)].\end{aligned}\right\} \tag{21}$$

In these variables we find the commutation rules to be

$$[a_\epsilon(k),\, a_{\epsilon'}^*(k')] = [b_\epsilon(k),\, b_{\epsilon'}^*(k')] = \delta_{\epsilon\epsilon'}\,\delta(k,k'), \tag{22}$$

while all $a(k)$ commute with all $b(k)$, etc. Further, we obtain for the Hamiltonian

$$H = \sum_k\sum_{\epsilon=-1}^{+1}\hbar c\sqrt{(k^2+\kappa^2)}\,[a_\epsilon^*(k)\,a_\epsilon(k)+b_\epsilon^*(k)\,b_\epsilon(k)+1)], \tag{23}$$

and for the total charge

$$\frac{1}{c}\int s^0\,dV = e\sum_k\sum_{\epsilon=-1}^{+1}[a_\epsilon^*(k)\,a_\epsilon(k) - b_\epsilon^*(k)\,b_\epsilon(k)]. \tag{24}$$

As a consequence of (22) the operators

$$N^+(k) = a^*(k)\,a(k) \quad\text{and}\quad N^-(k) = b^*(k)\,b(k)$$

have all positive integers and 0 as eigenvalues. Thus, taking account of the expressions (23) and (24), we can obviously interpret the N as numbers of particles present in the field, the N^+ denoting particles of positive charge, the N^- negatively charged ones. This result is identical with that of the scalar case, except for the additional suffix ϵ occurring here, which indicates the three possible values that the spin component in the k-direction can assume. As the N are not restricted to the values 0 and 1, the particles have Bose statistics.

The energy of a particle of the kind k is $E(k) = \hbar c\sqrt{(k^2+\kappa^2)}\ T$, its rest mass therefore $m_0 = \kappa\hbar/c$. It is of importance to note that, in spite of the formal analogy in the final result, the treatment of the "longitudinal" part of the field, which gives the quanta with $\epsilon = 0$, was necessarily separate from that of the "transversal" field which gives the other two spin orientations. The latter part corresponds exactly to the electrodynamical case, whereas there is no analogy to the former there.

Proca has already given the extension of his formulae for the case of the presence of electromagnetic fields. His equations then are

$$\left(\frac{\partial}{\partial x_\alpha}-\frac{ie}{\hbar c}A_\alpha\right)\phi_\beta-\left(\frac{\partial}{\partial x_\beta}-\frac{ie}{\hbar c}A_\beta\right)\phi_\alpha=\kappa\chi_{\alpha\beta}, \tag{25}$$

$$\left(\frac{\partial}{\partial x_\alpha}-\frac{ie}{\hbar c}A_\alpha\right)\chi^{\alpha\beta}=\kappa\phi^\beta, \tag{26}$$

the A_α being the electromagnetic potentials. Just as in the scalar case the addition of the potentials makes no difference to the procedure of quantization. The commutation rules between the ϕ_i and the χ^*_{0i} remain the same as before, and the expression for the charge density remains unaltered. The Hamiltonian acquires two additional terms, which are again very similar to those given by Pauli and Weisskopf in the scalar case:

$$\begin{aligned} H_1 = {}&-\frac{ie}{\hbar\kappa}\int dV\{\operatorname{curl}\Psi^*.[A\times(\Psi+\Pi^*)]+\operatorname{div}\Psi^*[A.(\Psi-\pi^*)] \\ &\qquad\qquad -\text{complex conjugate}\} \\ &-\frac{e^2}{\hbar^2\kappa c}\int dV\{|A\times(\psi+\Pi^*)|^2+|A.(\Psi-\pi^*)|^2\} \end{aligned} \tag{27}$$

and $$H_2=\frac{ie}{\hbar}\int dV A^0(\Psi-\pi^*).(\psi^*+\Pi)+\text{complex conjugate}. \tag{28}$$

(The symbol . means scalar multiplication whereas the vectorial product is denoted by ×.)

This formalism appears sufficiently simple to justify the attempt to connect it with experience. Though more general cases will be studied in the following, the final result will be that the case described in detail above is the one best suited to account for the facts of nuclear interaction.

For the rigorous derivation of the expression describing the interaction of Proca particles with neutrons and protons the considerations of § 3 are essential. It may, however, be useful to note that in applications the necessary extension consists solely in adding to the Hamiltonian the term given later as equation (63 *b*). This expression can be understood using the ideas of the present section alone. It gives rise to emission and absorption of heavy electrons by protons and neutrons.

3. GENERAL THEORY OF EINSTEIN-BOSE PARTICLES

In a recent paper Dirac (1936) has studied some wave equations for particles with a spin momentum *n* times that of the electron or proton.

Apart from the usual Dirac equation of the electron the simplest case contained in his scheme, for non-vanishing rest mass, is the following (we use spinor notation and take the case of no electromagnetic field only):

$$\left.\begin{aligned} p^{\dot{\alpha}\kappa}A_{\kappa\lambda} &= \sqrt{2}\; m_0 B^{\dot{\alpha}}_{\lambda}, \\ p_{\dot{\alpha}\kappa}B^{\dot{\alpha}}_{\lambda} &= \frac{m_0}{\sqrt{2}} A_{\kappa\lambda}. \end{aligned}\right\} \tag{29}$$

This case corresponds to the spin eigenvalue 1 (twice the electron spin). For Dirac's classification the assumption that $A_{\kappa\lambda}$ is symmetrical in its two suffixes is essential. Let us however compare the equations with those ensuing from the opposite assumption of antisymmetry. For an arbitrary antisymmetrical spinor we can raise one suffix according to the general rule and then put

$$A^{\lambda}_{\kappa} = A\delta^{\lambda}_{\kappa}, \tag{30}$$

where A is a scalar. Equations (29) then become

$$\left.\begin{aligned} p^{\dot{\alpha}}_{\kappa}A &= \sqrt{2}\;\; m_0 B^{\dot{\alpha}}_{\kappa}, \\ p^{\kappa}_{\dot{\alpha}}B^{\dot{\alpha}}_{\kappa} &= -\sqrt{2}\;\; m_0 A. \end{aligned}\right\} \tag{31}$$

It is not necessary to go into the details of spinor notation here, as we will now abandon it; it was merely introduced to obtain an initial classification of possible wave theories. The transition from equations (29) or (31) to tensor equations can be performed immediately; it is, however, not unique. In the spinor notation no specification of the transformation of the wave functions for reflexions of the co-ordinate system is implied. As a consequence, when reverting to tensor notation we can always choose either of two dual quantities to correspond to any spinor. Thus, one possibility of writing equations (31) in tensor form is

$$\frac{\partial\phi}{\partial x_\alpha} = \kappa\chi_\alpha, \tag{32a}$$

$$\frac{\partial\chi^\alpha}{\partial x_\alpha} = \kappa\phi, \tag{33a}$$

$$(\kappa = m_0 c/\hbar)$$

and similarly the first possibility for equations (29) is

$$\frac{\partial\phi_\beta}{\partial x_\alpha} - \frac{\partial\phi_\alpha}{\partial x_\beta} = \kappa\chi_{\alpha\beta}, \tag{32b}$$

$$\frac{\partial\chi^{\alpha\beta}}{\partial x_\alpha} = \kappa\phi^\beta. \tag{33b}$$

Alternatively, however, (29) can be transcribed as

$$\frac{\partial\phi_{\beta\gamma}}{\partial x_\alpha}-\frac{\partial\phi_{\alpha\gamma}}{\partial x_\beta}+\frac{\partial\phi_{\alpha\beta}}{\partial x_\gamma} = \kappa\chi_{\alpha\beta\gamma}, \tag{32c}$$

$$\frac{\partial\chi^{\alpha\beta\gamma}}{\partial x_\alpha} = \kappa\phi^{\beta\gamma}, \tag{33c}$$

$$(\phi_{\alpha\beta} = -\phi_{\beta\alpha})$$

and (31) as

$$\frac{\partial\phi_{\beta\gamma\delta}}{\partial x_\alpha}-\frac{\partial\phi_{\alpha\gamma\delta}}{\partial x_\beta}+\frac{\partial\phi_{\alpha\beta\delta}}{\partial x_\gamma}-\frac{\partial\phi_{\alpha\beta\gamma}}{\partial x_\delta} = \kappa\chi_{\alpha\beta\gamma\delta} \tag{32d}$$

$$\frac{\partial\chi^{\alpha\beta\gamma\delta}}{\partial x_\alpha} = \kappa\phi^{\beta\gamma\delta} \tag{33d}$$

$$(\phi_{\alpha\beta\gamma} = -\phi_{\beta\alpha\gamma} = -\phi_{\alpha\gamma\beta}).$$

Now, equations (32*a*) and (33*a*) are obviously equivalent to the scalar relativistic Schrödinger equation (Klein-Gordon equation)

$$\left(\sum_i \frac{\partial^2}{\partial x_i^2}-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\phi = \kappa^2\phi, \tag{34}$$

χ_α being merely a symbol for the 4-gradient of the wave function. Wave equation (34) also holds for all the other field quantities in the above four cases, but in general the statement of the second-order relation (34) alone is not sufficient.

(32*b*) and (33*b*) are Proca's equations. We have, however, further found two cases of obviously highly analogous structure and equal simplicity which we have not considered before. It is true, of course, that their difference from the two first cases is to a large extent a matter of notation, the quantity χ_{0123} in the last set, for instance, behaving like the scalar ϕ in case *a* in all but reflexions (with $\phi_{123} \leftrightarrow -\frac{1}{c}\phi, \phi_{0ik} \leftrightarrow -\mathrm{grad}_l\,\phi$). This difference in reflexion character does not in the slightest affect the quantization of the theory; in case *c* the method of the preceding section, in case *d* that of Pauli and Weisskopf is immediately applicable, and we obtain particles of spin 1 or zero just as before. Nevertheless, the separate treatment of the dual cases is necessary as soon as the terms describing the interaction of this field with protons and neutrons are introduced, as these are characteristically different in either of two dual cases. From now onwards we will therefore consider the four cases independently.

We begin by noting that the four sets of equations can each be derived from a Lagrangian, namely,

$$L_0^a = -c\kappa(\phi^*\phi + \chi_\alpha^*\chi^\alpha), \tag{35a}$$

$$L_0^b = -c\kappa(\phi_\alpha^*\phi^\alpha + \tfrac{1}{2}\chi_{\alpha\beta}^*\chi^{\alpha\beta}), \tag{35b}$$

$$L_0^c = -c\kappa(\tfrac{1}{2}\phi_{\alpha\beta}^*\phi^{\alpha\beta} + \tfrac{1}{6}\chi_{\alpha\beta\gamma}^*\chi^{\alpha\beta\gamma}), \tag{35c}$$

$$L_0^d = -c\kappa(\tfrac{1}{6}\phi_{\alpha\beta\gamma}^*\phi^{\alpha\beta\gamma} + \tfrac{1}{24}\chi_{\alpha\beta\gamma\delta}^*\chi^{\alpha\beta\gamma\delta}). \tag{35d}$$

Taking L^n as a function of the independent variables ϕ, and expressing the χ as derivatives of the ϕ by means of the equations ($32n$), the equations ($33n$) can be deduced as variational equations in the usual manner. We can further easily determine the complete expressions for the energy-momentum tensor and the current density vector:

$$T_{\alpha\beta}^a = (\chi_\alpha^*\chi_\beta + \text{complex conjugate}) - \delta_{\alpha\beta}L_0^a, \tag{36a}$$

$$T_{\alpha\beta}^b = (\phi_\alpha^*\phi_\beta + \chi_\alpha^{*\gamma}\chi_{\beta\gamma} + \text{complex conjugate}) - \delta_{\alpha\beta}L_0^b, \tag{36b}$$

$$T_{\alpha\beta}^c = (\phi_\alpha^{*\gamma}\phi_{\beta\gamma} + \tfrac{1}{2}\chi_\alpha^{*\gamma\delta}\chi_{\beta\gamma\delta} + \text{complex conjugate}) - \delta_{\alpha\beta}L_0^c, \tag{36c}$$

$$T_{\alpha\beta}^d = (\tfrac{1}{2}\phi_\alpha^{*\gamma\delta}\phi_{\beta\gamma\delta} + \tfrac{1}{6}\chi_\alpha^{*\gamma\delta\epsilon}\chi_{\beta\gamma\delta\epsilon} + \text{complex conjugate}) - \delta_{\alpha\beta}L_0^d, \tag{36d}$$

$$s_\alpha^a = \frac{eci}{\hbar}(\chi_\alpha^*\phi - \chi_\alpha\phi^*), \tag{37a}$$

$$s_\alpha^b = \frac{eci}{\hbar}(\chi_{\alpha\beta}^*\phi^\beta - \chi_{\alpha\beta}\phi^{\beta*}), \tag{37b}$$

$$s_\alpha^c = \frac{eci}{2\hbar}(\chi_{\alpha\beta\gamma}^*\phi^{\beta\gamma} - \chi_{\alpha\beta\gamma}\phi^{\beta\gamma*}), \tag{37c}$$

$$s_\alpha^d = \frac{eci}{6\hbar}(\chi_{\alpha\beta\gamma\delta}^*\phi^{\beta\gamma\delta} - \chi_{\alpha\beta\gamma\delta}\phi^{\beta\gamma\delta*}). \tag{37d}$$

In each case the energy density is essentially positive; this is a special characteristic of these simple cases. If we had proceeded from any more complicated examples of Dirac's spinor equations, we would not in general have obtained this result. There thus appears to be some justification for choosing these four cases of simple and similar type for preferential consideration, and it seems certain that any alternative formalism would have to be a great deal more complicated for physically satisfactory results to be expected. A theory for spin values greater than unity therefore immediately leads to serious difficulties. On the other hand, the strength of the analogy among the four possibilities here stated is so evident that it would not appear justified to exclude any of them *a priori*. For reasons of economy it is of course to be hoped that a comparison with experience will prove that one of the cases alone is sufficient to give a description of reality.

When one proceeds to the formulation of the law of interaction of the new particles with protons and neutrons one observes yet again that all four cases appear equally adequate for these purposes. The quantities occurring in our possible fundamental equations are one scalar (ϕ), two 4-vectors (χ_α and ϕ_α), two antisymmetrical tensors of the second order ($\chi_{\alpha\beta}$ and $\phi_{\alpha\beta}$), two "pseudovectors", that is, totally antisymmetrical tensors of the third order ($\chi_{\alpha\beta\gamma}$ and $\phi_{\alpha\beta\gamma}$), and finally one "pseudoscalar" (χ_{0123}). We now assume that the protons and neutrons obey the Dirac equation; then, from their wave functions, we can form just one covariant of each of these five types and no others. Denoting the Dirac wave function of the proton by Φ_P, that of the neutron by Φ_N, we can form the five covariants

$$\begin{array}{lll} \text{Scalar:} & \Phi_N^* \beta \Phi_P & = w, \\ \text{Vector:} & (-\Phi_N^* \Phi_P;\ \Phi_N^* \alpha_i \Phi_P) & = (v_0;\ v_i), \\ \text{Antisymmetrical tensor:} & (\Phi_N^* \gamma_i \Phi_P;\ \Phi_N^* \beta \sigma_l \Phi_P) & = (u_{0i};\ u_{ik}), \\ \text{Pseudovector:} & (\Phi_N^* \sigma_l \Phi_P;\ -\Phi_N^* \gamma_5 \Phi_P) & = (t_{0ik};\ t_{123}), \\ \text{Pseudoscalar:} & \Phi^* \beta \gamma_5 \Phi_P & = s_{0123}. \end{array} \tag{38}$$

α_i, β are the Dirac matrices, $\gamma_i = -i\beta\alpha_i$, $\sigma_l = -i\alpha_i\alpha_k$ and $\gamma_5 = -i\alpha_1\alpha_2\alpha_3$.

Further, we can of course form the complex conjugates of the above quantities, for instance,

$$\Phi_P^* \beta \Phi_N = w^*.$$

We will now put down the most general interaction satisfying the following conditions:

(i) The interaction term to be added to the Lagrangian involves only the wave functions of protons and neutrons without their derivatives.

(ii) The field quantities describing the heavy electron which enter the interaction expression are the ϕ and χ only, derivatives of these quantities not occurring explicitly.

In each of the four cases there are then two terms that may be added to the Lagrangian if the condition of relativistic invariance is to be satisfied. The first in each case involves the ϕ, namely:

$$L_g^a = -c\kappa(g_a w \phi^* + \text{conjugate complex}), \tag{39a}$$

$$L_g^b = -c\kappa(g_b v_\alpha \phi^{\alpha *} + \text{conjugate complex}), \tag{39b}$$

$$L_g^c = -c\kappa(g_c \tfrac{1}{2} u_{\alpha\beta} \phi^{\alpha\beta *} + \text{conjugate complex}), \tag{39c}$$

$$L_g^d = -c\kappa(g_d \tfrac{1}{6} t_{\alpha\beta\gamma} \phi^{\alpha\beta\gamma *} + \text{conjugate complex}), \tag{39d}$$

The other term will contain the χ:

$$L_f^a = -c\kappa(f_a v_\alpha \chi^{\alpha *} + \text{conjugate complex}), \tag{40a}$$

$$L_f^b = -c\kappa(f_b \tfrac{1}{2} u_{\alpha\beta} \chi^{\alpha\beta *} + \text{conjugate complex}), \tag{40b}$$

$$L_f^c = -c\kappa(f_c \tfrac{1}{6} t_{\alpha\beta\gamma} \chi^{\alpha\beta\gamma *} + \text{conjugate complex}), \tag{40c}$$

$$L_f^d = -c\kappa(f_d \tfrac{1}{24} s_{\alpha\beta\gamma\delta} \chi^{\alpha\beta\gamma\delta *} + \text{conjugate complex}). \tag{40d}$$

The f_n and g_n are arbitrary constants.†

The general interaction will of course be a linear combination of the two terms. The assumptions (i) and (ii) are justified on grounds of simplicity; it might, however, be here objected that the interactions from (40*n*) are already less simple than those given by (39*n*), as the χ are in fact, by (32*n*), derivatives of the "wave functions" ϕ. From the example treated in § 2, we have, however, already seen that the theory is quite symmetrical with respect to the ϕ and the χ, and that in fact, for the longitudinal field, it is natural to consider ϕ as defined by (33*n*), χ then being considered the "wave function". Besides, there is complete analogy between the ϕ of cases *a* and *b* and the χ of cases *d* and *c* respectively, and vice versa, and there is therefore no legitimate reason for considering the second kind of interaction to be less simple. Our present procedure of treating the ϕ as wave functions and the χ as their derivatives is necessary only for formal reasons.

The total Lagrangian can now be assumed to have the form

$$\bar{L} = \int dV[L_0^n + L_g^n + L_f^n + L_P + L_N], \tag{41}$$

where L_P and L_N are the Lagrangians of the free protons and neutrons respectively. These are of course well known, as we assume the Dirac equation to hold for the heavy particles, and we need not put them down explicitly.

Defining the canonical conjugate $p(q)$ of any co-ordinate q in the usual way,

$$p(q) = \frac{\partial \bar{L}}{\partial \dot{q}},$$

---

† The quantities defined as g and f in the paper by Fröhlich and others (1938) are related to g_b and f_b by

$$g = -\left(\frac{\kappa c}{4\pi}\right)^{\frac{1}{2}} g_b, \quad f = +\left(\frac{\kappa c}{4\pi}\right)^{\frac{1}{2}} f_b.$$

we obtain
$$p(\phi) = \zeta_0^* = \chi_0^* + f_a^* v_0^*, \tag{42a}$$

$$p(\phi_i) = \zeta_{0i}^* = \chi_{0i}^* + f_b^* u_{0i}^*, \tag{42b}$$

$$p(\phi_{ik}) = \zeta_{0ik}^* = \chi_{0ik}^* + f_c^* t_{0ik}^*, \tag{42c}$$

$$p(\phi_{123}) = \zeta_{0123}^* = \chi_{0123}^* + f_d^* s_{0123}^*, \tag{42d}$$

$$p(\phi_0) = p(\phi_{0i}) = p(\phi_{0ik}) = 0. \tag{43}$$

As there are no time derivatives of the proton or neutron wave functions in L_g^n or L_f^n, there are no additions to the canonical conjugates of the heavy particle wave functions due to the interaction.

We can now pass to the Hamiltonian function by the usual procedure:

$$H = \overline{\Sigma p\dot{q}} - \bar{L}.$$

We take case b as an example:

$$\begin{aligned} H^b = c\kappa \int dV \Big[& \phi_i^* \phi_i + \phi_0^* \phi_0 + \tfrac{1}{2}\chi_{ik}^* \chi_{ik} + \zeta_{0i}^* \zeta_{0i} \\ & + (g_b v_i \phi_i^* + f_b \tfrac{1}{2} u_{ik} \chi_{ik}^* - f_b u_{0i} \zeta_{0i}^* + \text{conjugate complex}) \\ & + |f_b|^2 u_{0i}^* u_{0i} \\ & - \left\{ \phi_0^* \left(\phi_0 + v_0 g - \frac{1}{\kappa} \frac{\partial}{\partial x_i} \zeta_{i0} \right) + \text{conjugate complex} \right\} \Big]. \end{aligned} \tag{44}$$

The final term is derived by means of a partial integration. This form corresponds to the final form of the Hamiltonian of electrodynamics. However, here again, as in the simple case studied in § 2, we can consider the equation

$$\frac{\partial \zeta_{i0}}{\partial x_i} = \kappa(\phi_0 + v_0 g), \tag{45}$$

not as a supplementary condition, but as the definition of ϕ_0, and we proceed to eliminate this quantity from the Hamiltonian. H is then a function of ϕ_i and ζ_{0i} only. Whereas such a consideration is not necessary in case a, it reappears in an almost identical form in cases c and d. There we will use the abbreviations

$$\kappa\eta_0 = -\frac{\partial \zeta_{0i}}{\partial x_i}, \quad \kappa\eta_{0i} = +\frac{\partial \zeta_{0ik}}{\partial x_k}, \quad \kappa\eta_{0ik} = -\frac{\partial \zeta_{0ikl}}{\partial x_l}. \tag{46}$$

The four Hamiltonians can then be expressed as

$$H^n = H_0^n + H_1^n + H_2^n, \tag{47}$$

$$H_0^a = c\kappa \int dV[\phi^*\phi + \zeta_0^*\zeta_0 + \chi_i^*\chi_i], \tag{48a}$$

$$H_0^b = c\kappa \int dV[\phi_i^*\phi_i + \eta_0^*\eta_0 + \zeta_{0i}^*\zeta_{0i} + \tfrac{1}{2}\chi_{ik}^*\chi_{ik}], \tag{48b}$$

$$H_0^c = c\kappa \int dV[\tfrac{1}{2}\phi_{ik}^*\phi_{ik} + \eta_{0i}^*\eta_{0i} + \tfrac{1}{2}\zeta_{0ik}^*\zeta_{0ik} + \tfrac{1}{6}\chi_{ikl}^*\chi_{ikl}], \tag{48c}$$

$$H^d = c\kappa \int dV[\tfrac{1}{6}\phi_{ikl}^*\phi_{ikl} + \tfrac{1}{2}\eta_{0ik}^*\eta_{0ik} + \tfrac{1}{6}\zeta_{0ikl}^*\zeta_{0ikl}], \tag{48d}$$

$$H_1^a = c\kappa \int dV[g_a w\phi^* - f_a(v_0\zeta_0^* - v_i\chi_i^*) + \text{conjugate complex}], \tag{49a}$$

$$H_1^b = c\kappa \int dV[g_b(v_i\phi_i^* - v_0\eta_0^*) - f_b(u_{0i}\zeta_{0i}^* - \tfrac{1}{2}u_{ik}\chi_{ik}^*) + \text{conjugate complex}], \tag{49b}$$

$$H_1^c = c\kappa \int dV[g_c(\tfrac{1}{2}u_{ik}\phi_{ik}^* - u_{0i}\eta_{0i}^*) - f_c(\tfrac{1}{2}t_{0ik}\zeta_{0ik}^* - \tfrac{1}{6}t_{ikl}\chi_{ikl}) \\ + \text{conjugate complex}], \tag{49c}$$

$$H_1^d = c\kappa \int dV[g_d(\tfrac{1}{6}t_{ikl}\phi_{ikl}^* - \tfrac{1}{2}t_{0ik}\eta_{0ik}^*) - f_d\tfrac{1}{6}s_{0ikl}\zeta_{0ikl}^* + \text{conjugate complex}], \tag{49d}$$

$$H_2^a = |f_a|^2 v_0^* v_0, \tag{50a}$$

$$H_2^b = |g_b|^2 v_0^* v_0 + |f_b|^2 u_{0i}^* u_{0i}, \tag{50b}$$

$$H_2^c = |g_c|^2 u_{0i}^* u_{0i} + |f_c|^2 \tfrac{1}{2}t_{0ik}^* t_{0ik}, \tag{50c}$$

$$H_2^d = |g_d|^2 \tfrac{1}{2}t_{0ik}^* t_{0ik} + |f_d|^2 \tfrac{1}{6}s_{0ikl}^* s_{0ikl}. \tag{50d}$$

The Hamiltonians of the free heavy particles have been omitted. To the Hamiltonians we must adjoin the definitions

$$\zeta_i - f_a v_i = \chi_i = \frac{1}{\kappa}\frac{\partial\phi}{\partial x_i}; \tag{51a}$$

$$\zeta_{ik} - f_b u_{ik} = \chi_{ik} = \frac{1}{\kappa}\left(\frac{\partial\phi_k}{\partial x_i} - \frac{\partial\phi_i}{\partial x_k}\right); \quad \phi_0 + g_b v_0 = \eta_0 = -\frac{1}{\kappa}\frac{\partial\zeta_{0i}}{\partial x_i}; \tag{51b}$$

$$\zeta_{ikl} - f_c t_{ikl} = \chi_{ikl} = \frac{1}{\kappa}\left(\frac{\partial\phi_{kl}}{\partial x_i} - \frac{\partial\phi_{il}}{\partial x_k} + \frac{\partial\phi_{ik}}{\partial x_l}\right); \quad \phi_{0i} + g_c u_{0i} = \eta_{0i} = +\frac{1}{\kappa}\frac{\partial\zeta_{0ik}}{\partial x_k}; \tag{51c}$$

$$\phi_{0ik} + g_d u_{0ik} = \eta_{0ik} = -\frac{1}{\kappa}\frac{\partial\zeta_{0ikl}}{\partial x_l}; \tag{51d}$$

and then, in the usual way, we obtain the canonical equations

$$\frac{1}{\kappa}\frac{\partial\phi}{\partial x_0} = \zeta_0 - f_a v_0; \quad \frac{1}{\kappa}\left(-\frac{\partial\zeta_0}{\partial x_0}+\frac{\partial\zeta_i}{\partial x_i}\right) = \phi + g_a w; \tag{52a}$$

$$\frac{1}{\kappa}\left(\frac{\partial\phi_i}{\partial x_0}-\frac{\partial\phi_0}{\partial x_i}\right) = \zeta_{0i} - f_b u_{0i}; \quad \frac{1}{\kappa}\left(-\frac{\partial\zeta_{0i}}{\partial x_0}+\frac{\partial\zeta_{ki}}{\partial x_k}\right) = \phi_i + g_b v_i; \tag{52b}$$

$$\frac{1}{\kappa}\left(\frac{\partial\phi_{ik}}{\partial x_0}-\frac{\partial\phi_{0k}}{\partial x_i}+\frac{\partial\phi_{0i}}{\partial x_k}\right) = \zeta_{0ik} - f_c t_{0ik}; \quad \frac{1}{\kappa}\left(-\frac{\partial\zeta_{0ik}}{\partial x_0}+\frac{\partial\zeta_{lik}}{\partial x_l}\right) = \phi_{ik} + g_c u_{ik}; \tag{52c}$$

$$\frac{1}{\kappa}\left(\frac{\partial\phi_{ikl}}{\partial x_0}-\frac{\partial\phi_{0kl}}{\partial x_i}+\frac{\partial\phi_{0il}}{\partial x_k}-\frac{\partial\phi_{0ik}}{\partial x_l}\right) = \zeta_{0ikl} - f_d s_{0ikl}; \quad \frac{1}{\kappa}\left(\frac{\partial\zeta_{0ikl}}{\partial x_0}\right) = \phi_{ikl} + g_d t_{ikl}. \tag{52d}$$

The relativistic invariance of the scheme is evident.

In cases *c* and *d* the present notation is awkward. For any antisymmetrical tensor let us by definition put

$$\bar{a}_0 = a_{ik}, \quad \bar{a}_{ik} = -a_{0l}, \tag{53}$$

for any pseudovector $\bar{b}_0 = b_{123}, \quad \bar{b}_l = b_{0ik},$ (54)

and for a pseudoscalar $\bar{c} = c_{0123}.$ (55)

The expressions (48) to (50) for the Hamiltonians then become

$$H_0^c = c\kappa\int dV[\bar{\zeta}_i^*\bar{\zeta}_i + \bar{\chi}_0^*\bar{\chi}_0 + \bar{\phi}_{0i}^*\bar{\phi}_{0i} + \tfrac{1}{2}\bar{\eta}_{ik}^*\bar{\eta}_{ik}], \tag{56c}$$

$$H_0^d = c\kappa\int dV[\bar{\zeta}^*\bar{\zeta} + \bar{\phi}_0^*\bar{\phi}_0 + \bar{\eta}_i^*\bar{\eta}_i], \tag{56d}$$

$$H_1^c = -c\kappa\int dV[f_c(\bar{t}_i\bar{\zeta}_i^* - \bar{t}_0\bar{\chi}_0^*) - g_c(\bar{u}_{0i}\bar{\phi}_{0i}^* - \tfrac{1}{2}\bar{u}_{ik}\bar{\eta}_{ik}^*) + \text{conjugate complex}], \tag{57c}$$

$$H_1^d = -c\kappa\int dV[f_d\bar{s}\bar{\zeta}^* - g_d(\bar{t}_0\bar{\phi}_0^* - \bar{t}_i\bar{\eta}_i^*) + \text{conjugate complex}], \tag{57d}$$

$$H_2^c = |f_c|^2\bar{t}_i^*\bar{t}_i + |g_c|^2\tfrac{1}{2}\bar{u}_{ik}^*\bar{u}_{ik}, \tag{58c}$$

$$H_2^d = |f_d|^2\bar{s}^*\bar{s} + |g_d|^2\bar{t}_i^*\bar{t}_i. \tag{58d}$$

The form of (48*a*, *b*)–(50*a*, *b*) is very similar to (56*c*, *d*)–(58*c*, *d*) respectively, there being a correspondence between the following sets of quantities

$$\phi\leftrightarrow\bar{\zeta}, \quad \eta\leftrightarrow\bar{\chi}, \quad \zeta\leftrightarrow\bar{\phi}, \quad \chi\leftrightarrow\bar{\eta},$$

$$g_c\leftrightarrow -f_b, \quad g_d\leftrightarrow -f_a, \quad f_c\leftrightarrow -g_b, \quad f_d\leftrightarrow -g_a.$$

However, this correspondence is not quite complete. Instead of the terms (58*c*) and (58*d*) actually found one would, from analogy with (50*b*) and (50*a*), have expected

$$H_2'^c = |f_c|^2 \bar{t}_0^* \bar{t}_0 + |g_c|^2 \bar{u}_{0i}^* \bar{u}_{0i}, \tag{59c}$$

and

$$H_2'^d = |g_d|^2 \bar{t}_0^* \bar{t}_0. \tag{59d}$$

This is the form of the terms quadratic in the f_n or g_n which we would have obtained if we had neglected to treat the dual cases separately but had put down the results for them from analogy. The reason for the difference is however very simple. In our original Lagrangian we assumed the ϕ to be the independent variables, or in our new notation for cases *c* and *d* the $\bar{\phi}'$. These however correspond to the χ of cases *a* and *b*. In these first two cases we would indeed have had final terms of the form

$$H_2'^a = |g_a|^2 w^* w + |f_a|^2 v_i^* v_i \tag{60a}$$

and

$$H_2'^b = |g_b|^2 v_i^* v_i + |f_b|^2 \tfrac{1}{2} u_{ik}^* u_{ik} \tag{60b}$$

instead of (50*a*) and (50*b*), if we had adopted the following procedure:

As Lagrangian we take

$$\bar{L} = \int dV[-L_0^n - L_g^n - L_f^n + L_P + L_N]. \tag{61}$$

We then express the ϕ as derivatives of the χ by means of equations (33*n*), and take the χ as the quantities that are varied independently. This method is equivalent to the one actually chosen for the corresponding dual cases. There thus exists an ambiguity in the formulation of the theory which would appear surprising, and there seems to be no reason for preferring either of the two forms. Fortunately the ambiguity is not serious. The undetermined terms H_2^n or $H_2'^n$, it is important to note, do not involve the wave functions of the heavy electron but only those of proton and neutron. They can be interpreted as a direct interaction between these particles with the δ-function as potential. Precisely these interactions have been studied in a previous paper (Kemmer 1937), and it was seen that they do not give any finite binding energy between the particles concerned. Similar considerations can also be employed to show that this type of interaction gives no scattering cross section whatever.† The non-vanishing contributions due to such terms are solely of the self-energy type. For such questions the theory is, however, in any case inadequate, so that the additional ambiguity introduced here is of no consequence at all. It might even be hoped that one could utilize the indeterminateness here found to reduce the divergence of self energy calculations, but the final section of this paper will show that such an attempt cannot succeed.

---

† This result was also proved very generally in unpublished calculations by Professor Pauli, who gave the writer his kind permission to quote the results.

For the present it is certainly allowable to ignore the terms H_2^n and $H_2'^n$ completely.

For quantization let us now revert to vector notation as in § 2. We can then put

$$H_0^a = H_0^d = c\kappa\int dV\left[\,|\psi|^2 + |\pi|^2 + \frac{1}{\kappa^2}|\operatorname{grad}\psi|^2\right], \tag{62 a, d}$$

$$\begin{aligned} H_0^b = H_0^c = c\kappa\int dV\Big[\,|\vec{\psi}|^2 + |\vec{\pi}|^2 + \frac{1}{\kappa^2}|\operatorname{curl}\psi|^2 \\ + |\vec{\Psi}|^2 + |\vec{\Pi}|^2 + \frac{1}{\kappa^2}|\operatorname{div}\Psi|^2\Big], \end{aligned} \tag{62 b, c}$$

$$\begin{aligned} H_1^a = c\kappa\int dV\Big\{\Phi_N^*\Big[g_a\beta\psi^* - f_a\Big(\pi - \frac{1}{\kappa}(\vec{\alpha}\operatorname{grad}\psi^*)\Big)\Big]\Phi_P \\ + \text{conjugate complex}\Big\}, \end{aligned} \tag{63a}$$

$$\begin{aligned} H_1^b = c\kappa\int dV\Big\{\Phi_N^*\Big[g_b(\vec{\alpha}\,\vec{\psi}^*) - f_b\Big((\vec{\gamma}\,\vec{\pi}) - \frac{1}{\kappa}\beta(\vec{\sigma}\operatorname{curl}\psi^*)\Big) + f_b(\vec{\gamma}\,\vec{\Psi}^*) \\ + g_b\Big((\vec{\alpha}\,\vec{\Pi}) - \frac{1}{\kappa}\operatorname{div}\Psi^*\Big)\Big]\Phi_P + \text{conjugate complex}\Big\}, \end{aligned} \tag{63b}$$

$$\begin{aligned} H_1^c = -c\kappa\int dV\Big\{\Phi_N^*\Big[f_c(\vec{\sigma}\,\vec{\psi}^*) - g_c\Big(\beta(\vec{\sigma}\,\vec{\pi}) + \frac{1}{\kappa}(\vec{\gamma}\operatorname{curl}\psi^*)\Big) + g_c\beta(\vec{\sigma}\,\vec{\Psi}^*) \\ + f_c\Big((\vec{\sigma}\,\vec{\Pi}) - \frac{1}{\kappa}\gamma_5\operatorname{div}\Psi^*\Big)\Big]\Phi_P + \text{conjugate complex}\Big\}. \end{aligned} \tag{63c}$$

$$\begin{aligned} H_1^d = -c\kappa\int dV\Big\{\Phi_N^*\Big[f_d\beta\gamma_5\psi^* - g_d\Big(\gamma_5\pi - \frac{1}{\kappa}(\vec{\sigma}\operatorname{grad}\psi^*)\Big)\Big]\Phi_P \\ + \text{conjugate complex}\Big\}. \end{aligned} \tag{63d}$$

Here the ψ, Ψ and π, Π are understood to be *defined* as canonical conjugates. The whole of the further development can now be taken over without change from the case of no interaction.† In particular the equations from (12) to (24) in § 2, or the corresponding equations of Pauli and Weisskopf, remain valid. It is of special importance that the total charge will still be, in cases a and d:

$$\frac{1}{c}\int s_0\,dV = \frac{ei}{\hbar}\int(\pi\psi - \pi^*\psi^*)\,dV, \tag{64}$$

and in cases b and c:

$$\frac{1}{c}\int s_0\,dV = \frac{ei}{\hbar}\int(\vec{\pi}\vec{\psi} + \vec{\Pi}\vec{\Psi} - \vec{\pi}^*\vec{\psi}^* - \vec{\Pi}^*\vec{\Psi}^*)\,dV. \tag{65}$$

The terms (63n) describe possibility of the emission and absorption of heavy electrons by protons and neutrons, and therefore the heavy electron charge alone will of course not be conserved. It can however be shown that if one adds the usual Dirac charge $\int\Phi_P^*\Phi_P\,dV$ of the protons to the quantity defined in (64) or (65), their sum is indeed a constant. As, on the other hand, $\int s_0\,dV$ is the same operator as in the case of no interaction, the interpretation of the field theory in terms of numbers of charges present will still be rigorously possible.

4. PERTURBATION THEORY; THE NEUTRON-PROTON INTERACTION

As is well known, the applications of formalisms of the type sketched above are in practice confined to perturbation theory. We will first deal with the most important problem of this kind, namely the derivation of the neutron-proton interaction from our field theories. This will be given by the second order perturbation formula

$$\begin{aligned}(p'_P p'_N \,|\, V(r) \,|\, p_N p_P) = &-\Sigma\frac{(p'_P p'_N \,|\, H_1^n \,|\, k p_N p'_N)(k p_N p'_N \,|\, H_1^n \,|\, p_N p_P)}{E'_N - E_P + \epsilon^+} \\ &-\Sigma\frac{(p'_P p'_N \,|\, H_1^n \,|\, k p'_P p_P)(k p'_P p_P \,|\, H_1^n \,|\, p_N p_P)}{E'_P - E_N + \epsilon^-}.\end{aligned} \tag{66}$$

---

† The equations can immediately be generalized to the case when an electromagnetic field is present. One has merely to perform the usual substitutions

$$\frac{\partial\psi}{\partial x_i} \to \left(\frac{\partial}{\partial x_i} - \frac{ie}{\hbar c}A_i\right)\psi, \text{ etc.}$$

The resulting addition to (62b, c) has already been given as equation (27). It should, however, not be overlooked that these are also additions to the H_1^n of equations (63n).

E_N and E_P are the energies corresponding to the momenta p_N and p_P of neutron and proton respectively; $\hbar k$ denotes the momentum of the heavy electron, ϵ its energy, and the sum in (66) has to be taken over all values of this momentum. The first part of (66) takes into account the transitions of the proton into a neutron state under emission of a positive heavy electron, which is subsequently reabsorbed by the neutron, the latter thereby becoming a proton. In the latter case the neutron is the emitting, the proton the absorbing particle and the heavy electron is of negative charge. The two terms give equal contributions. If one only considers heavy particles at distances $r \gg h/Mc$ (M being the proton mass), the recoil of the heavy particle in the heavy electron emissions may be neglected and the energy difference in the denominators of (66) can be put equal to the heavy electron energy ϵ. With this assumption we find the following expressions for the effective neutron-proton potential $V(r)$:†

$$V^a(r) = +\frac{c\kappa}{4\pi}\left[-g_a^2\beta_N\beta_P + f_a^2\frac{1}{\kappa^2}(\vec{\alpha}_N\,\mathrm{grad})\,(\vec{\alpha}_P\,\mathrm{grad})\right]J(r), \tag{67 a}$$

$$\begin{aligned} V^b(r) = +\frac{c\kappa}{4\pi}\Bigg[g_b^2\Big(1-(\vec{\alpha}_N\vec{\alpha}_P)+\frac{1}{\kappa^2}(\vec{\alpha}_N\,\mathrm{grad})\,(\vec{\alpha}_P\,\mathrm{grad})\Big)\\ +f_b^2\Big(\beta_N\beta_P(\vec{\sigma}_N\vec{\sigma}_P)-\frac{1}{\kappa^2}\beta_N\beta_P(\vec{\sigma}_N\,\mathrm{grad})\,(\vec{\sigma}_P\,\mathrm{grad})\\ -\frac{1}{\kappa^2}(\vec{\gamma}_N\,\mathrm{grad})\,(\vec{\gamma}_P\,\mathrm{grad})\Big)\Bigg]J(r), \end{aligned} \tag{67 b}$$

$$\begin{aligned} V^c(r) = +\frac{c\kappa}{4\pi}\Bigg[f_c^2\Big(\gamma_{5N}\gamma_{5P}-(\vec{\sigma}_N\vec{\sigma}_P)+\frac{1}{\kappa^2}(\vec{\sigma}_N\,\mathrm{grad})\,(\vec{\sigma}_P\,\mathrm{grad})\Big)\\ +g_c^2\Big((\vec{\gamma}_N\vec{\gamma}_P)-\frac{1}{\kappa^2}(\vec{\gamma}_N\,\mathrm{grad})\,(\vec{\gamma}_P\,\mathrm{grad})\\ -\frac{1}{\kappa^2}\beta_N\beta_P(\vec{\sigma}_N\,\mathrm{grad})\,(\vec{\sigma}_P\,\mathrm{grad})\Big)\Bigg]J(r), \end{aligned} \tag{67 c}$$

$$V^d(r) = +\frac{c\kappa}{4\pi}\left[-f_d^2\beta_N\beta_P\gamma_{5N}\gamma_{5P} + g_d^2\frac{1}{\kappa^2}(\vec{\sigma}_N\,\mathrm{grad})\,(\vec{\sigma}_P\,\mathrm{grad})\right]J(r), \tag{67 d}$$

$$J(r) = \frac{2}{\pi r}\int_0^\infty \frac{k\sin kr\,dk}{k^2+\kappa^2}. \tag{68}$$

---

† For the purposes of this and the following sections the possibility of taking the g_n and f_n to be complex quantities is inessential. For brevity we will treat them as real.

In obtaining equations (67n) certain integrals of the type

$$\frac{1}{r}\int kr\sin kr\,dr$$

were neglected. For $r \neq 0$ they are indeed zero, but to be exact they should be represented by $-\frac{1}{r}\delta'(r)$. However, to these the same considerations apply as were already used in the preceding section; their sole contributions are to the self energies of the heavy particles. Therefore their omission is obviously justified.

The value of the integral $J(r)$ in equations (68) can be immediately given, we have

$$J(r) = -\frac{2}{\pi r}\frac{d}{dr}\int_0^\infty \frac{\cos kr}{k^2+\kappa^2}\,dk = \frac{e^{-\kappa r}}{r}.$$

Having neglected the recoil of the protons and neutrons, it is now not very consequent to use the relativistic terms of equations (67). It is more sensible to neglect the contributions proportional to the heavy particle velocities, that is the terms with α or γ, the same is true for the terms with γ_5. Further we should put $\beta = 1$.

Thus we obtain the final formulae:

$$V^a(r) = -\frac{c\kappa}{4\pi}g_a^2\frac{e^{-\kappa r}}{r}, \tag{69a}$$

$$V^b(r) = +\frac{c\kappa}{4\pi}\left[g_b^2 + f_b^2\left((\vec{\sigma}_N\vec{\sigma}_P) - (\vec{\sigma}_N\,\mathrm{grad})(\vec{\sigma}_P\,\mathrm{grad})\right)\right]\frac{e^{-\kappa r}}{r}, \tag{69b}$$

$$V^c(r) = -\frac{c\kappa}{4\pi}\Big[g_c^2(\vec{\sigma}_N\,\mathrm{grad})(\vec{\sigma}_P\,\mathrm{grad}) + f_c^2\Big((\vec{\sigma}_N\vec{\sigma}_P) - (\vec{\sigma}_N\,\mathrm{grad})(\vec{\sigma}_P\,\mathrm{grad})\Big)\Big]\frac{e^{-\kappa r}}{r}, \tag{69c}$$

$$V^d(r) = +\frac{c\kappa}{4\pi}g_d^2(\vec{\sigma}_N\,\mathrm{grad})(\vec{\sigma}_P\,\mathrm{grad})\frac{e^{-\kappa r}}{r}. \tag{69d}$$

Taking the most general form of our theory that is possible, namely a linear combination of all four cases, we see at once that we get an expression

$$V(r) = [A + B(\vec{\sigma}_N\vec{\sigma}_P) + C(\vec{\sigma}_N\,\mathrm{grad})(\vec{\sigma}_P\,\mathrm{grad})]\frac{e^{-\kappa r}}{r}, \tag{70}$$

with completely arbitrary coefficients A, B and C. This result has been stated before (Kemmer 1938). In this way a form of interaction with $C = 0$

may of course also be chosen, but though the latter assumption has been common to most calculations on nuclear binding, there appears to be no reason why it should be maintained. It is certainly more satisfactory to abandon it, if in this way one could succeed in describing all empirical facts with one of the four possibilities a to d alone. *This is actually possible using case b.* The detailed discussion of this fact is the subject of the paper of Fröhlich and others. Using the same methods as there described, we will here only briefly state the results for all four cases, which justify the preference for case b.

Case a: The potential in the ^{3}S ground state of the deuteron is proportional to g_a^2, and thus repulsive. The potential in the ^{1}S state is of equal magnitude, but attractive. Interaction f_a^2 gives no non-relativistic contribution.

Case b: The potential in the ^{3}S state is $\sim -(g_b^2 + \frac{2}{3} f_b^2)$ and thus always attractive. The potential in the ^{1}S state is $\sim g_b^2 - 2f_b^2$ and so can be chosen to be about half as deep as that of the triplet state, as experiment requires.

Case c: The potential in the ^{3}S state, $\sim \frac{2}{3} f_c^2 + \frac{1}{3} g^2$, is repulsive, there is a repulsion 3 times as strong in the ^{1}S state.

Case d: The potential in the ^{3}S state is attractive $(-\frac{1}{3} g_d^2)$, but there is a larger attraction $(-g_d^2)$ in the singlet state. The interaction f_d^2 gives no non-relativistic contribution.

It is obvious that no other case but b agrees with experience. To achieve agreement with experiment the inclusion of the interaction f_b is of course essential.† We have stressed the symmetry of our theory with respect to the ϕ and the χ sufficiently to make it clear that it cannot be excluded by arguments of simplicity.

Some remarks on perturbation theory in general are of interest in this connexion. If we consider the interaction expression (63b), we note that there are terms proportional to $g_b \psi^*$ and other terms $\sim \frac{1}{\kappa} g_b \frac{\partial}{\partial x} \Psi^*$ combined in the same invariant, the former arising from the transversal part, the latter from the longitudinal. The transversal part is similar to the electrodynamical case. Therefore the evaluation of succeeding approximations of an effect by perturbation theory will for this term be in powers of an absolute constant, corresponding to $(e^2/\hbar c)^{\frac{1}{2}}$ in radiation theory. In our choice of units this constant is $g_b(\kappa/4\pi\hbar)^{\frac{1}{2}}$. The contributions due to the additional longitudinal term, on the other hand, will be a power series in $g_b(1/4\pi\hbar\kappa)^{\frac{1}{2}}$, that is, we have an expansion in powers of a fundamental length. Using the present field theory the two contributions will always occur together. This is the case even for $f_b = 0$, and the same is true for any interaction in case *c*. The state of affairs is different in cases *a* and *d*, as there the interactions g_a and f_d give terms of the first type only.

---

† The case $f_b = 0$ has been considered by Bhabha (1938 a).

Now Heisenberg (1936) has pointed out that, in a theory involving a fundamental length in the sense stated above, the contribution of high order approximations will increase when the energies in the effects studied increase, and thus multiple processes such as cosmic ray showers should, according to such a theory, occur when sufficiently high energies are available. This is not the case for formalisms involving expansions in powers of a dimensionless constant only. Unpublished calculations by Pauli have also shown that in such a theory the well-known divergency difficulties become mathematically far more serious than in radiation theory. It seems therefore of importance to note that the data on neutron-proton interaction prove definitely that a theory of the simpler type is not sufficient, and that according to the views developed here Heisenberg showers should be expected to occur.

It must be further remarked that quite generally the expansions as performed in perturbation theory can naturally only be valid if the expansion parameters are small. As is shown in the paper of Fröhlich and others by the evaluation of the 4th order proton-proton force, this is here by no means the case. However, it is well known that the question of avoiding these expansions is the main unsolved difficulty of quantum field theory, and therefore the use of any better method does not at present appear practicable. It seems fairly certain, that even with better methods the result that a general linear combination of our possibilities can be made to fit known data would be maintained, but of course the statement that *b* alone is sufficient, is more open to criticism. The study of the cases other than *b* may therefore still prove to be of more than systematic value.

5. SELF-ENERGIES

Hitherto we have avoided the consideration of effects which, calculated by our theory, give divergent results. It is well known that all forms of quantum field theory formulated up to now give such divergencies, as soon as the calculations involve the process of virtual emission and reabsorption of a quantum by one and the same particle. The self-energy of a proton or a neutron due to the interaction with our heavy electrons is an effect of this type. We do not intend to discuss here whether one should attach any physical significance to these calculations in spite of their formal divergence, but it is in any case of importance to be informed as to the exact type of these divergencies in simple cases. We will therefore give a brief summary of the results of such calculations.

For these questions it is not in general sensible to maintain the assumption of the last section that the recoil (energy change) of the heavy particles in the processes considered can be neglected; therefore the explicit assumption is now necessary that protons and neutrons obey Dirac's equation including the "hole theory" interpretation of negative energy states. The methods of evaluating self energies in "hole theory" may be found in the paper by Weisskopf (1934*a*, *b*), where the case of electromagnetic interaction is considered.

Here we will have two types of self energies to evaluate, namely those due to the interactions H_1^n and those arising from the additions H_2^n or $H_2'^n$, which, of course, can now no more be neglected. The former terms give a self energy in second order perturbation, namely

$$W_1^n = -\sum_{\substack{p \\ (E>0)}} \frac{(p_0 \,|\, H_1^n \,|\, kp)\,(kp \,|\, H_1^n \,|\, p_0)}{E - E_0 + \epsilon} + \sum_{\substack{p \\ (E<0)}} \frac{(p_0 \,|\, H_1^n \,|\, kp)\,(kp \,|\, H_1^n \,|\, p_0)}{|\,E\,| + E_0 + \epsilon}, \quad (71)$$

the notation being the same as in the previous section. p_0 and E_0 are the momentum and energy of the proton (or neutron) in its initial state, p and E the corresponding quantities in the intermediate state, and $\hbar k$ and ϵ the same for the heavy electron. The general expression for W_1^n is found to be not very simple. We will confine ourselves to giving the result under two alternative special assumptions only:

(1) We take the proton to be at rest: $p_0 = 0$.

Then with the designation M for the mass of the proton and $K = Mc/\hbar$, we have

$$W_1^a = \int dk \{g_a^2 \kappa^2 A(k) + f_a^2 [(k^2 + \kappa^2)\, A(k) + k^2 B(k)]\}, \qquad (72a)$$

$$W_1^b = \int dk (g_b^2 + f_b^2)\, [k^2 A(k) + (k^2 + 3\kappa^2)\, B(k)], \qquad (72b)$$

$$W_1^c = \int dk (g_c^2 + f_c^2)\, [(k^2 + 3\kappa^2)\, A(k) + k^2 B(k)], \qquad (72c)$$

$$W_1^d = \int dk \{g_d^2 [k^2 A(k) + (k^2 + \kappa^2)\, B(k)] + f_d^2 \kappa^2 B(k)\}, \qquad (72d)$$

with

$$\left.\begin{aligned} A(k) &= -\frac{c}{4\pi^2\kappa}\frac{k^2}{\sqrt{(k^2+K^2)}\sqrt{(k^2+\kappa^2)}}\left[\frac{\sqrt{(k^2+K^2)}+K}{\sqrt{(k^2+\kappa^2)}+\sqrt{(k^2+K^2)}-K}\right. \\ &\qquad\left.-\frac{\sqrt{(k^2+K^2)}-K}{\sqrt{(k^2+\kappa^2)}+\sqrt{(k^2+K^2)}+K}\right], \\ B(k) &= +\frac{c}{4\pi^2\kappa}\frac{k^2}{\sqrt{(k^2+K^2)}}\left[\frac{1}{\sqrt{(k^2+\kappa^2)}+\sqrt{(k^2+K^2)}-K}\right. \\ &\qquad\left.-\frac{1}{\sqrt{(k^2+\kappa^2)}+\sqrt{(k^2+K^2)}+K}\right]. \end{aligned}\right\} \tag{73}$$

(2) We admit an arbitrary value for p_0, but give only the first term of the expansion in descending powers of k:

$$\left.\begin{aligned} &W_1^a = f_a^2 C_1, \quad W_1^b = (g_b^2+2f_b^2)\,C_1, \\ &W_1^c = (2g_c^2+f_c^2)\,C_1, \quad W_1^d = g_d^2 C_1, \\ &C_1 = -\frac{1}{2\pi^2\hbar\kappa}\frac{M^2+\frac{1}{3}p_0^2}{E_0}\int dk\,k + \dots. \end{aligned}\right\} \tag{74}$$

Thus the divergence of the result is quadratic in every case, but cases a and d do not include such highly divergent terms if we put $f_a = 0$, or respectively $g_d = 0$. In these special cases the divergence can be shown to be logarithmical only. It should further be noted that the representation (74) is only valid for $k \gg K$. Thus if one "cuts off" the integrals at k_0 and, as is done by Fröhlich and others, takes $k_0 \sim \kappa$, expression (74) cannot be used. Taking (73) instead we then find

$$\left.\begin{aligned} &W_1^a = f_a^2 D, \quad W_1^b = (g_b^2+f_b^2)\,D, \\ &W_1^c = (g_c^2+f_c^2)\,D, \quad W_1^d = g_d^2 D, \\ &D = -\frac{c}{2\pi^2\kappa}\int dk\,k^2 + \dots, \end{aligned}\right\} \tag{75}$$

these being the highest terms of an expansion in powers of K/k_0.

We now determine the contributions of the H_2^n and $H_2'^n$. These are simpler to calculate, as one needs only first-order perturbation theory:

$$W_2^n = \sum_{E>0}(p_0 p \mid H_2^n \mid p_0 p) - \sum_{E<0}(p_0 p \mid H_2^n \mid p_0 p), \tag{76}$$

and after a short calculation

$$\left.\begin{aligned} W_2^a &= f_a^2 C_2, & W_2'^a &= (g_a^2 - 3f_a^2)\, C_2, \\ W_2^b &= (g_b^2 - 3f_b^2)\, C_2, & W_2'^b &= -3(g_b^2 - f_b^2)\, C_2, \\ W_2^c &= -3(g_c^2 - f_c^2)\, C_2, & W_2'^c &= (3g_c^2 - f_c^2)\, C_2, \\ W_2^d &= (3g_d^2 - f_d^2)\, C_2, & W_2'^d &= -g_d^2 C_2, \\ C_2 &= \frac{1}{\pi^2 \hbar \kappa} \frac{M^2}{E_0} \int \frac{dk\, k^2}{\sqrt{(k^2 + K^2)}}. \end{aligned}\right\} \tag{77}$$

Here, too, the degree of divergence is quadratic. The dependence on p_0 is, however, different in these terms, and it is therefore evident that they will in no case compensate the other divergencies, even if we make use of the arbitrariness in the choice of the H_2^n.

The fact that the divergence of self-energy is here greater than in radiation theory is noteworthy. Case b might especially have been expected to yield results identical with those for electromagnetic interaction, at least for $f_b = 0$. However, the contribution of the longitudinal quanta is sufficient to alter the result essentially. This is obviously in direct connexion with the facts concerning Heisenberg showers discussed in the foregoing section, as is particularly well shown by the fact that just those interactions, which do not give showers (g_a and f_d) also give no quadratically divergent contributions.

A further quantity that must strictly speaking be classed among the divergent effects is the magnetic moment of the proton or neutron caused by the possibility of emitting heavy electrons. This has been studied for case b in the paper by Fröhlich and others, but as the derivation there is by an abridged method, we would here like to indicate the more rigorous method of calculating the effect. In § 2 we gave the expressions for the interaction of the heavy electrons with the electromagnetic field. It was given there for case 2 only, but it is very similar in all four cases. Denoting the operator there named H_1 by $H_R(A)$, we can obtain a third order perturbation self energy dependent on the electromagnetic potentials (recoil being neglected):

$$\begin{aligned} W(A) = &-\Sigma \frac{(p_0 \mid H_1^n \mid k'p)(k'p \mid H_R(A) \mid kp)(kp \mid H_1^n \mid p_0)}{k'k} \\ &-\Sigma \frac{(p_0 \mid H_1^n \mid k^+ p)(k^+ p \mid H_1^n \mid k^+ k^- p_0)(k^+ k^- p_0 \mid H_R(A) \mid p_0)}{k^+(k^+ + k^-)} \\ &-\Sigma \frac{(p_0 \mid H_R(A) \mid k^+ k^- p_0)(k^+ k^- p_0 \mid H_1^n \mid k^+ p)(k^+ p \mid H_1^n \mid p_0)}{(k^+ + k^-)\, k^+}. \end{aligned} \tag{78}$$

The first term describes the virtual emission of a positive heavy electron by the proton, its subsequent interaction with the electromagnetic field and its final reabsorption by the proton. This term is identical with the expression used in the other paper. The second term describes another effect of the same order, namely the creation of a pair of heavy electrons by the field and their consecutive absorption by the proton. Term 3 describes the process inverse to this one. For the evaluation of the magnetic moment we need only take a static magnetic field, which can be described by a time-independent vector potential A satisfying the condition

$$\operatorname{div} A = 0. \tag{79}$$

The magnetic moment is then defined as

$$\mu = \left(\frac{\partial W(A)}{\partial H}\right)_{H=0}, \tag{80}$$

$H = \operatorname{curl} A$ being the magnetic field strength; to obtain the term proportional to H in (78) we must expand in powers of $k-k'$ or k^+-k^-, i.e. the Fourier component of the magnetic field, and restrict ourselves to the linear term of this expansion.

Using relation (79) it is easy to prove that

(1) In the expansion of the denominators one need only use the zero-order term.

(2) Only the terms in the heavy electron interactions proportional to σ_P or $\beta_P\sigma_P$ can give a contribution to the moment (for a proton at rest).

From (1) it follows that the second and third terms of (78) give just the contribution of the first term again, so that the method of calculation used in the other paper is reasonable; from (2) we deduce that the only interactions giving rise to a moment are f_b, g_c, f_c and g_d.

However doubtful the details of these calculations may be considered in view of the divergence of the result, the fact that an interaction has to be preferred which indicates the existence of an additional moment must be considered as satisfactory. The degree of divergence is here in all cases linear (see Fröhlich and others 1938).

The writer would like to express his thanks to Professor W. Pauli and to Professor G. Wentzel of Zürich for the interest taken in this work and the valuable suggestions given, to Professor Pauli also for the permission to refer to unpublished investigations.

SUMMARY

It is shown that there are four inequivalent but equally simple possibilities of formulating a field theory of Einstein-Bose particles, in which a positive expression for the energy density exists. Any of these formalisms might tentatively be accepted as a description of the "heavy electron". Considerations of relativistic invariance show that two independent expressions for the interaction of these particles with protons and neutrons can be chosen in each of the four cases. Taking account of the interaction terms the general Hamiltonian form of the theories is stated and the quantization is performed. The resulting proton-neutron potential is determined and it is found that its sign and spin-dependence agrees with reality in only one of the four cases, namely in the case based on the equations of Proca (1936). The (divergent) self-energies of the proton or neutron resulting from the interactions studied are evaluated.

REFERENCES

Bhabha 1938*a* *Nature, Lond.*, **141**, 117.
Bhabha 1938*b* *Proc. Roy. Soc.* A (in the Press).
Dirac 1936 *Proc. Roy. Soc.* A, **155**, 447.
Durandin and Erschow 1937 *Phys. Z. Sowjet.* **12**, 466.
Fröhlich and Heitler 1938 *Nature, Lond.*, **141**, 37.
Fröhlich, Heitler and Kemmer 1938 *Proc. Roy. Soc.* A, **166**, 154.
Heisenberg 1932 *Z. Phys.* **77**, 1.
— 1936 *Z. Phys.* **101**, 533.
Kemmer 1937 *Helv. phys. Acta*, **10**, 47.
— 1938 *Nature, Lond.*, **141**, 116.
Neddermeyer and Anderson 1937 *Phys. Rev.* **51**, 884.
Oppenheimer and Serber 1937 *Phys. Rev.* **51**, 1113.
Pauli and Weisskopf 1934 *Helv. phys. Acta*, **7**, 709.
Proca 1936 *J. Phys. Radium* (vii), **7**, 347.
Stueckelberg 1937 *Phys. Rev.* **52**, 41.
Weisskopf 1934*a* *Z. Phys.* **89**, 27.
— 1934*b* *Z. Phys.* **90**, 817.
Yukawa 1935 *Proc. Phys. Math. Soc. Japan*, **17**, 48.
— 1937 *Proc. Phys. Math. Soc. Japan*, **19**, 712.
Yukawa and Sakata 1937 *Proc. Phys. Math. Soc. Japan*, **19**, 1084.

On the nuclear forces and the magnetic moments of the neutron and the proton

By H. Fröhlich, *H. H. Wills Physical Laboratory, University of Bristol,*

W. Heitler, *Research Fellow of the Academic Assistance Council, H. H. Wills Physical Laboratory, Bristol,*

and N. Kemmer, *Beit Scientific Research Fellow, Imperial College, London*

(*Communicated by N. F. Mott, F.R.S.—Received* 1 *February* 1938)

Introduction

It was first suggested by Heisenberg that the forces between a proton and a neutron are connected with an exchange of charge between the two heavy particles. This exchange nature of the neutron-proton forces is now generally accepted. It would follow from this assumption that in suitable circumstances a proton (neutron) could emit a positively (negatively) charged particle transforming itself into a neutron (proton). At first sight it seemed that the emission of positive or negative electrons in the β-decay could in this way be made responsible for the nuclear forces. This was, in fact, suggested by Iwanenko (1934) and Tamm (1934). It has also been pointed out by Wick (1935) that the virtual emission of β-electrons might explain the values of the magnetic moments of the proton and the neutron. These theories, however, were not successful. The nuclear forces, for instance, turn out to be too small by a factor of more than 10^{10} and have far too small a range; this is due to the fact that the β-decay constant is extremely small. Since the β-decay is a process which, in nuclear dimensions, takes "geological ages", one might think that the ordinary properties of the heavy particles have no direct connexion with this process and that an approximate theory of the nuclear forces should be possible without the inclusion of the β-decay.

A new hope for such an "exchange theory" of the properties of nuclei is offered by the probable existence of a hitherto unknown type of particle constituting the hard component of cosmic radiation. Since these particles do not lose much energy by radiation, it has been suggested by Neddermeyer and Anderson (1937) that they are (positive and negative) "heavy electrons" with a mass between that of an electron and a proton. From cosmic-ray data the mass of these particles can hardly be determined yet, but it can be limited to values between 3 and 300 electron masses. There are, however, some arguments favouring a mass nearer to the upper limit of 100–200 electron masses.†

The heavy electrons are certainly not stable. They seem to be absorbed strongly as soon as they reach an energy of less than 200×10^6 e-volts. It is probable that this absorption is due to some sort of *nuclear* processes. It seems, therefore, plausible to assume that a heavy electron can be absorbed directly by a proton or neutron. Thus we shall propose the existence of the following processes

$$\left.\begin{aligned} P &\rightleftarrows N + Y_+, \\ N &\rightleftarrows P + Y_-, \end{aligned}\right\} \tag{1}$$

where Y_+, Y_- denotes a positive and negative heavy electron. There is no need for the introduction of any neutrino. It follows from (1) that the heavy electrons have zero (or an integral) spin and satisfy Bose statistics.

It has already been shown by Yukawa (1935) that processes of the type (1) can account for the nuclear forces in a reasonable way. The range of the forces $1/\lambda$ is directly connected with the mass m_0 of the heavy electron: $1/\lambda \sim \hbar/m_0 c$. This leads to a value of $m_0 \sim 100$ electron masses. On the other hand, it will be shown in this paper that the processes (1) can also be made responsible for the magnetic moments of the neutron and proton. Independently of any detailed theory, it is possible to determine m_0 from the observed values of the magnetic moments (cf. Fröhlich and Heitler 1938). It turns out that m_0 is also about 100 electron masses.

The fact that m_0 can be determined in two independent ways both leading to the same value seems to us to give strong support for such an exchange theory of the nuclear properties based on assumption (1).

We think, therefore, that it might be a reasonable policy to try to link up the nuclear properties (forces and magnetic moments) with the cosmic-ray phenomena of the hard component rather than with the β-decay. As a first step in this direction, we shall try in this paper to build up a consistent scheme for the wave field of the heavy electrons and their interaction with the nuclear particles. It will be seen that such a scheme is possible and that it leads to a consistent theory of the nuclear forces and of the magnetic moments.

---

† According to the theory the ionization of a fast charged particle passing through a gas has a minimum for energies of the order of the rest energy of the particle and increases logarithmically for higher energies. The cosmic-ray experiments, however, show that the ionization of particles of the order of 10^9 e-volts is not appreciably higher than that of particles of 10^7–10^8 e-volts. The particles with energy $> 3 \times 10^8$ e-volts are practically all heavy electrons. If the rest mass of the latter were only a few electron masses the ionization should be nearly the same as for electrons, which is contrary to the experiments. The above result suggests therefore a mass of the order of at least 100 electron masses.

1. THE WAVE EQUATIONS FOR THE HEAVY ELECTRON AND ITS INTERACTION WITH NUCLEAR PARTICLES

Our first task is to find the wave equation for the heavy electron. The simplest relativistic wave equation for particles satisfying Bose statistics is the Klein-Gordon equation for a scalar wave function ψ in the quantized form of Pauli and Weisskopf (1934). This scheme has been used recently by Yukawa and Sakata (1937) to calculate the nuclear properties. It turns out, however, that such a theory cannot account for the magnetic moment of the proton and neutron and leads to a wrong spin dependence of the nuclear forces.

A scalar wave function is not the only possible one for Bose particles. Light quanta, for instance, are described by a *vector field*. It has been shown by Proca (1936) that such a vector field can also be used to describe Bose particles with a finite rest mass.

In this paper we shall assume a *vector wave function for the heavy electron* satisfying the equations given by Proca. In an accompanying paper Kemmer (1938) has studied a number of possible alternatives including both the case of a scalar wave function and the vector case. Only for the vector formalism the sign and spin dependence of the neutron-proton force is in agreement with empirical evidence. The other possibilities can therefore be disregarded here.

The quantization of Proca's field equations can be carried out by the method developed by Pauli and Weisskopf for a scalar wave function. The only difference is that a heavy electron can exist in three different states of polarization. For a particle with given momentum there are two waves with transverse polarization and—in contrast to the Maxwell case—one wave with longitudinal polarization.

An interesting feature of Proca's equations is their similarity to Maxwell's equations. Owing to this analogy the quantization may be performed in close correspondence to the well-known procedure in radiation theory. In the following we use a notation similar to that of Heitler's (1936*a*) book on radiation theory (cf. Chapter I, § 6). One slight formal difference must, however, be accepted from the start, due to the fact that the wave functions for the heavy electron are essentially complex quantities, whereas the radiation theory deals with real field variables. This results in a doubling of the number of variables; to each quantity there exists the complex conjugate.

Let ϕ_α be the wave function of the heavy electron; ϕ_α corresponds to the 4-vector potential in radiation theory. We use Greek suffices (from 1 to 4) to denote a 4-vector and Latin suffices (from 1 to 3) for a 3-vector. Instead of ϕ_4 we also introduce $\phi_0 = -i\phi_4$; ϕ_0 corresponds to the scalar potential.

We now define an antisymmetrical tensor corresponding to the field strengths of electrodynamics:

$$\chi_{\alpha\beta} = \frac{\partial \phi_\beta}{\partial x_\alpha} - \frac{\partial \phi_\alpha}{\partial x_\beta}. \tag{2}$$

For the complex conjugate quantities we have

$$\chi^*_{\alpha\beta} = \frac{\partial \phi^*_\beta}{\partial x_\alpha} - \frac{\partial \phi^*_\alpha}{\partial x_\beta}.$$

The wave equations in the absence of an external field are then the following:

$$\sum_\alpha \frac{\partial \chi_{\alpha\beta}}{\partial x_\alpha} = \lambda^2 \phi_\beta, \quad \lambda = \frac{m_0 c}{\hbar}. \tag{3}$$

A similar equation holds for $\chi^*_{\alpha\beta}$, ϕ^*_β. m_0 is the rest mass of the heavy electron. For $\lambda = 0$ the equations (2) and (3) are identical with Maxwell's equations.

From (2) and (3) it follows that

$$\sum_\alpha \frac{\partial^2 \phi_\beta}{\partial x_\alpha^2} = \lambda^2 \phi_\beta \tag{4}$$

and also

$$\sum_\beta \frac{\partial \phi_\beta}{\partial x_\beta} = 0. \tag{5}$$

Thus, each wave function ϕ_β satisfies the Klein-Gordon equation (4). Equation (5) corresponds to the "Lorentz condition" in Maxwell's theory. It is here a *consequence* of the field equations (2), (3). This is not the case in Maxwell's theory where (5) has to be assumed as an additional condition (it plays there only the role of a convenient normalization of the vector potential).

It can easily be shown that an adequate expression for the energy density of our system is given by

$$W = \frac{1}{4\pi}\left\{\sum_{i=1}^{3} \chi^*_{0i}\chi_{0i} + \tfrac{1}{2}\sum_{i,k=1}^{3} \chi^*_{ik}\chi_{ik} + \lambda^2\left(\phi^*_0\phi_0 + \sum_{i=1}^{3}\phi^*_i\phi_i\right)\right\}. \tag{6}$$

The space part ϕ_i of the wave function can be divided into a transverse (I) and a longitudinal (II) part:

$$\phi_i = \phi_i^{\mathrm{I}} + \phi_i^{\mathrm{II}} \tag{7}$$

with

$$\operatorname{div} \phi_i^{\mathrm{I}} = 0$$

$$\operatorname{curl} \phi_i^{\mathrm{II}} = 0.$$

We treat the two parts separately.

The transverse part is exactly analogous to the transverse waves in the radiation theory. We expand ϕ_i^{I} in a series of plane waves; let $\mathbf{A}_r(\mathbf{r})$ be a complete set of orthogonal eigenfunctions with the condition of periodicity in a cube of volume 1:

$$\mathbf{A}_\rho = \sqrt{(4\pi c^2)}\, \mathbf{j}_\rho e^{i(\mathbf{k}_\rho \mathbf{r})} \tag{8}$$

the $\mathbf{j}_\rho$ being either of the two independent unit vectors perpendicular to the wave vector $\mathbf{k}_\rho$. Putting now

$$\left.\begin{aligned} \phi_i^{\mathrm{I}} &= \sum_\rho q_\rho(t)\, \mathbf{A}_\rho(\mathbf{r}), \\ \phi_i^{\mathrm{I}*} &= \sum_\rho q_\rho^*(t)\, \mathbf{A}_\rho^*(\mathbf{r}), \end{aligned}\right\} \tag{9}$$

the q_ρ and q_ρ^* satisfy the equations

$$\left.\begin{aligned} \ddot{q}_\rho + \nu_\rho^2 q_\rho &= 0, \\ \ddot{q}_\rho^* + \nu_\rho^2 q_\rho^* &= 0, \end{aligned}\right\} \tag{10}$$

$$\frac{1}{c^2}\nu_\rho^2 = k_\rho^2 + \lambda^2. \tag{11}$$

The three components of the wave vector k_ρ can assume all values which are integral multiples of 2π. Using the definitions

$$\dot{q}_\rho = p_\rho^*, \quad \dot{q}_\rho^* = p_\rho, \tag{12}$$

equations (10) can obviously be regarded as the canonical equations of the Hamiltonian

$$H^{\mathrm{I}} = \sum_\rho H_\rho^{\mathrm{I}} = \sum_\rho (p_\rho^* p_\rho + \nu_\rho^2 q_\rho^* q_\rho).$$

This expression can also be seen to be identical with the transverse part of the space integral of the energy density W given by equation (6).

The longitudinal wave functions ϕ_i^{II} and ϕ_0 may similarly be expanded in the form

$$\left.\begin{aligned} \phi_0 &= \sum_\sigma a_\sigma \phi_\sigma, \\ \phi_i^{\mathrm{II}} &= \sum_\sigma q_\sigma \mathbf{B}_\sigma, \quad (\operatorname{curl} \mathbf{B}_\sigma = 0) \end{aligned}\right\} \tag{13}$$

The ϕ_σ are a complete set of scalar wave functions and $\mathbf{B}_\sigma = (1/k_\sigma) \operatorname{grad} \phi_\sigma$ (cf. Heitler 1936*a*, pp. 47–55). The quantities q_σ, a_σ and their complex conjugates satisfy the same differential equations (10) as the q_ρ. If we put, therefore,

$$\dot{q}_\sigma = p_\sigma^*, \quad \dot{q}_\sigma^* = p_\sigma; \quad \dot{a}_\sigma = -b_\sigma^*, \quad \dot{a}_\sigma^* = -b_\sigma, \tag{14}$$

we may describe the system by means of the Hamiltonian

$$H^{\mathrm{II}} = \sum_\sigma [(p_\sigma^* p_\sigma + \nu_\sigma^2 q_\sigma^* q_\sigma) - (b_\sigma^* b_\sigma + \nu_\sigma^2 a_\sigma^* a_\sigma)] \tag{15}$$

(the minus sign of the second term is explained below). The quantities a_σ, q_σ, b_σ, p_σ are not independent of each other. From (5) it follows that

$$ck_\sigma q_\sigma = -b_\sigma^*, \tag{16}$$

and therefore

$$ck_\sigma p_\sigma = -\dot{b}_\sigma = -\nu_\sigma^2 a_\sigma^*. \tag{16'}$$

Inserting these equations into the Hamiltonian (15), H^{II} can be expressed by the q_σ, p_σ only:

$$H^{\mathrm{II}} = \sum_\sigma \frac{c^2\lambda^2}{\nu_\sigma^2} (p_\sigma^* p_\sigma + \nu_\sigma^2 q_\sigma^* q_\sigma). \tag{17}$$

(15) or (17) is again identical with the space integral of the longitudinal part of the energy density defined by (6). For this purpose it was essential to choose the a_σ, b_σ part of the Hamiltonian (15) with the minus sign (cf. also Heitler 1936*a*, p. 55).

The fact that the longitudinal part H^{II} does not vanish is a characteristic feature of our theory. In Maxwell's case ($\lambda = 0$) H^{II} vanishes except if charged particles are present and can then be reduced to a static Coulomb energy only. In our theory there are *longitudinal waves even in vacuo.*

The expression (17) has not exactly the form of an energy of an oscillator. But we can easily introduce new variables so as to re-establish this form. We consider the longitudinal part of the quantity χ_{0i} and define a wave function

$$\lambda\psi_i \equiv -\chi_{0i}^{\mathrm{II}} = \frac{\partial\phi_0}{\partial x_i} + \frac{1}{c}\phi_i. \tag{18}$$

Using the same $\mathbf{B}_\sigma$ as above, we can expand

$$\psi_i = \sum_\sigma Q_\sigma \mathbf{B}_\sigma, \tag{19}$$

and putting

$$\dot{Q}_\sigma = P_\sigma^*, \quad \dot{Q}_\sigma^* = P_\sigma \tag{20}$$

we find

$$\left.\begin{aligned} P_\sigma &= -c\lambda q_\sigma^*. \\ Q_\sigma &= \frac{c\lambda}{\nu_\sigma^2} p_\sigma^*. \end{aligned}\right\} \tag{21}$$

Hence

$$H^{\text{II}} = \sum_\sigma (P_\sigma^* P_\sigma + \nu_\sigma^2 Q_\sigma^* Q_\sigma). \tag{22}$$

Thus the longitudinal part of the field ϕ_0, ϕ_i^{II} is equivalent to *one* set of oscillators with the Hamiltonian (22). Thus the total Hamiltonian of our system is the sum for the transverse and longitudinal oscillators

$$H = \sum_\rho (p_\rho^* p_\rho + \nu_\rho^2 q_\rho^* q_\rho) + \sum_\sigma (P_\sigma^* P_\sigma + \nu_\sigma^2 Q_\sigma^* Q_\sigma). \tag{23}$$

The *quantization* of the field equations is now straightforward. We put

$$\left.\begin{aligned} p_\rho^* q_\rho^* - q_\rho^* p_\rho^* &= p_\rho q_\rho - q_\rho p_\rho = -i\hbar, \\ P_\sigma^* Q_\sigma^* - Q_\sigma^* P_\sigma^* &= P_\sigma Q_\sigma - Q_\sigma P_\sigma = -i\hbar, \end{aligned}\right\} \tag{24}$$

and assume all other pairs of quantities to commute.

Following the procedure of Pauli and Weisskopf (1934) we further put

$$\left.\begin{aligned} p_\rho &= \sqrt{\left(\frac{\epsilon_\rho}{2}\right)}(a_\rho^* + b_\rho), \quad p_\rho^* = \sqrt{\left(\frac{\epsilon_\rho}{2}\right)}(a_\rho + b_\rho^*), \\ q_\rho &= \frac{i\hbar}{\sqrt{(2\epsilon_\rho)}}(a_\rho - b_\rho^*), \quad q_\rho^* = \frac{-i\hbar}{\sqrt{(2\epsilon_\rho)}}(a_\rho^* - b_\rho), \\ P_\sigma &= \sqrt{\left(\frac{\epsilon_\sigma}{2}\right)}(A_\sigma^* + B_\sigma), \quad P_\sigma^* = \sqrt{\left(\frac{\epsilon_\sigma}{2}\right)}(A_\sigma + B_\sigma^*), \\ Q_\sigma &= \frac{i\hbar}{\sqrt{(2\epsilon_\sigma)}}(A_\sigma - B_\sigma^*), \quad Q_\sigma^* = \frac{-i\hbar}{\sqrt{(2\epsilon_\sigma)}}(A_\sigma^* - B_\sigma), \end{aligned}\right\} \tag{25}$$

where $\epsilon_\rho = \hbar\nu_\rho$ is the energy of the oscillator ρ.

Then the total energy becomes

$$H = \sum_\rho \epsilon_\rho (a_\rho^* a_\rho + b_\rho^* b_\rho + 1) + \sum_\sigma \epsilon_\sigma (A_\sigma^* A_\sigma + B_\sigma^* B_\sigma + 1) \tag{26}$$

and

$$\left.\begin{aligned} a_\rho^* a_\rho - a_\rho a_\rho^* &= b_\rho^* b_\rho - b_\rho b_\rho^* = -1, \\ A_\sigma^* A_\sigma - A_\sigma A_\sigma^* &= B_\sigma^* B_\sigma - B_\sigma B_\sigma^* = -1. \end{aligned}\right\} \tag{27}$$

From (27) it follows that the quantities

$$\left.\begin{aligned} a_\rho^* a_\rho = N_\rho^+, \quad b_\rho^* b_\rho = N_\rho^-, \\ A_\sigma^* A_\sigma = M_\sigma^+, \quad B_\sigma^* B_\sigma = M_\sigma^-, \end{aligned}\right\} \tag{28}$$

all have the eigenvalues 0, 1, 2, ...

We can therefore interpret $N_\rho^\pm$ as the number of "transverse heavy electrons" with charge $\pm e$ in the state ρ and $M_\sigma^\pm$ the number of "longitudinal heavy electrons" with charge $\pm e$ in the state σ.

The operators a_ρ, a_ρ^*, b_ρ, b_ρ^*, etc., have only matrix elements for transitions in which the number of positive (negative) electrons increases or decreases by one

$$\left.\begin{aligned} a_\rho^*(N_\rho^+ \to N_\rho^+ + 1) &= \sqrt{(N_\rho^+ + 1)}, \\ b_\rho^*(N_\rho^- \to N_\rho^- + 1) &= \sqrt{(N_\rho^- + 1)}, \\ a_\rho(N_\rho^+ \to N_\rho^+ - 1) &= \sqrt{N_\rho^+}, \\ b_\rho(N_\rho^- \to N_\rho^- - 1) &= \sqrt{(N_\rho^-)}. \end{aligned}\right\} \tag{29}$$

Corresponding equations hold for the $A_\sigma, \ldots$ and M_σ^+, M_σ^-. All other matrix elements vanish.

Our next task is to find an *interaction between the heavy electrons and the nuclear particles proton and neutron* leading to processes of the type (1). These processes are analogous to the emission of light by an electron or the emission of an electron and neutrino in the β-decay. To find the interaction in question we let ourselves be guided by the theory of light and Fermi's theory of the β-decay. In this paper we shall deal only with protons and neutrons moving very slowly compared with the velocity of light. We can therefore assume that the heavy particles are at rest at a given position $\mathbf{r}_0$, say.

We then assume:

(1) The interaction energy H' shall depend only on the wave function of the heavy electron (and its derivative) at the position of the heavy particle.

(2) It shall not depend on any variables of the heavy particles other than the spin vector $\boldsymbol{\sigma}$.

(3) It shall not contain higher order derivatives of the wave function of the heavy electron.

From relativistic arguments Kemmer (1938) has already deduced all possible interactions H' satisfying the above conditions (cf. equation (63 *b*) of his paper).[†] For our purpose we have simply to take the nonrelativistic parts of this interaction.

[†] The field equations (2)–(5) and the interaction (30 *b*) have also been obtained by Bhabha (1938).

There are two independent interactions for the transverse and longitudinal waves:

$$H'_{\mathrm{I}} = \frac{f}{\lambda}[\Pi(\boldsymbol{\sigma}\,\mathrm{curl}\,\phi_i) + \Pi^*(\boldsymbol{\sigma}\,\mathrm{curl}\,\phi_i)^*]_{(0)}, \tag{30a}$$

$$H'_{\mathrm{II}} = \frac{g}{\lambda}[\Pi\,\mathrm{div}\,\psi_i + \Pi^*\,\mathrm{div}\,\psi_i^*]_{(0)}, \tag{30b}$$

$$H' \;= H'_{\mathrm{I}} + H'_{\mathrm{II}}. \tag{30c}$$

Here Π is the operator transforming a neutron into a proton and is zero when applied to a proton. Π^* is the operator transforming a proton into a neutron and is zero when applied to a neutron. ϕ_i, ψ_i are the wave functions of the heavy electron for the transverse and longitudinal waves respectively (cf. equations (9) and (18)). g and f are two *constants* both with *dimensions of a charge*. We shall determine them later from experimental data. The index "0" in (30) indicates that the expressions in brackets have to be taken at the position of the heavy particle. $\boldsymbol{\sigma}$ is the Pauli spin vector:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{31}$$

The interaction (30) leads obviously to processes of the type (1). If we expand ϕ_i and ψ_i in the form (9) (19) we see from (25) and (29) that ϕ_i or ψ_i gives rise to an emission of a negative heavy electron or to an absorption of a positive heavy electron and this is, according to (30), connected with the transmutation of a neutron into a proton. The parts proportional to Π^* give rise to the inverse processes.

For later purposes, we finally give the matrix-elements of the interaction H' for the emission or absorption of a heavy electron represented by a plane wave $\mathbf{j}\exp i(\mathbf{kr})$ by a nuclear particle at the position $\mathbf{r}_0$. Taking into account equations (8), (9), (19), (25), (29), we find for

$$\left.\begin{aligned}
&\text{Emission of a longitudinal } Y^+ \text{ by proton: } && -\Pi^*\frac{g}{\lambda}\sqrt{\left(\frac{2\pi\hbar^2c^2}{\epsilon}\right)}e^{-i(\mathbf{kr}_0)}.k,\\
&\text{Emission of a longitudinal } Y^- \text{ by neutron: } && \Pi\,\frac{g}{\lambda}\sqrt{\left(\frac{2\pi\hbar^2c^2}{\epsilon}\right)}e^{i(\mathbf{kr}_0)}.k,\\
&\text{Emission of a transverse } Y^+ \text{ by proton: } && \Pi^*\frac{f}{\lambda}\sqrt{\left(\frac{2\pi\hbar^2c^2}{\epsilon}\right)}e^{-i(\mathbf{kr}_0)}(\boldsymbol{\sigma}.[\mathbf{jk}]),\\
&\text{Emission of a transverse } Y^- \text{ by neutron: } && -\Pi\,\frac{f}{\lambda}\sqrt{\left(\frac{2\pi\hbar^2c^2}{\epsilon}\right)}e^{i(\mathbf{kr}_0)}(\boldsymbol{\sigma}.[\mathbf{jk}]).
\end{aligned}\right\} \tag{32}$$

The matrix elements for absorption are the complex conjugates to (32).

In the following sections we apply this theory to the physical properties of the proton and neutron.

2. THE NEUTRON-PROTON FORCE

We first calculate the interaction energy of a proton and a neutron. The two particles are coupled with each other by the mutual emission and reabsorption of a heavy electron. This is accompanied by an exchange of charge between the proton and neutron and the force will, therefore, be an "exchange force". It follows from the perturbation theory that the operator of the interaction energy is given by

$$V_{NP} = -\sum_n \frac{H'_{An} H'_{nF}}{E_n - E_A}.$$

Here A is the state where a proton is found at the position $\mathbf{r}_1$ and a neutron at the position $\mathbf{r}_2$. F is the state where the proton is at $\mathbf{r}_2$ and the neutron at $\mathbf{r}_1$. n are all possible intermediate states in which a positive or negative heavy electron is emitted by one of the nuclear particles. H'_{An}, etc., are the matrix elements of H' given by equation (30). If we take for the intermediate states of the heavy electron plane waves with wave number $\mathbf{k}$, the matrix elements are given by (32). V_{NP} includes contributions from longitudinal and from transverse positive heavy electrons emitted by the proton (and absorbed by the neutron) and also from negative heavy electrons emitted by the neutron. The latter contribution is the same as that of the positive heavy electrons and results in a factor two.

The energy difference $E_n - E_A$ is just equal to the energy of the heavy electron

$$\epsilon = \hbar c \sqrt{(k^2 + \lambda^2)}. \tag{33}$$

Denoting the operator $\Pi^*(1)\,\Pi(2)$ which represents an exchange of charge between the two heavy particles by P, we obtain for the longitudinal contribution of V_{NP}

$$V_{\rm II} = -2\frac{g^2}{\lambda^2} 2\pi\hbar^2 c^2 P \Sigma \frac{k^2 e^{i(\mathbf{kr})}}{\epsilon^2}, \tag{34}$$

where the sum Σ has to be extended over all values of $\mathbf{k}$. $\mathbf{r}$ is the distance of the two heavy particles $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$. The sum can be replaced by an integral taking into account that the number of states with the direction of $\mathbf{k}$ lying in the element of the solid angle $d\Omega$ and k in the element dk is $d\Omega\, k^2 dk/(2\pi)^3$. Replacing k^2 by $(k^2+\lambda^2)-\lambda^2$ we obtain, besides an integral $\int \exp i(\mathbf{kr})\, d\mathbf{k} = 0$,

$$V_{\mathrm{II}} = +2\frac{g^2}{(2\pi)^2} P \int \frac{e^{i(\mathbf{kr})}\, d\Omega\, k^2 dk}{k^2+\lambda^2} = +g^2 P \frac{e^{-\lambda r}}{r}. \tag{35}$$

For the contribution of the transverse waves we obtain in a similar way

$$V_{\mathrm{I}} = -2\frac{f^2}{\lambda^2} 2\pi\hbar^2 c^2 P \Sigma \frac{(\boldsymbol{\sigma}[\mathbf{jk}])\,(\boldsymbol{\rho}[\mathbf{jk}])\, e^{i(\mathbf{kr})}}{\varepsilon^2}, \tag{36}$$

where $\boldsymbol{\sigma}$, $\boldsymbol{\rho}$ are the spin vectors of the particles 1 and 2 respectively. The Σ has also to be extended over all directions of polarization $\mathbf{j}$. Carrying out the latter summation we obtain

$$V_{\mathrm{I}} = -2\frac{f^2}{\lambda^2}\frac{P}{(2\pi)^2}\int \frac{k^4 dk\, d\Omega}{k^2+\lambda^2} e^{i(\mathbf{kr})} \left[(\boldsymbol{\sigma\rho}) - \frac{(\boldsymbol{\sigma}\mathbf{k})\,(\boldsymbol{\rho}\mathbf{k})}{k^2} \right].$$

Working out the integral in the same way as above, we obtain finally

$$V_{\mathrm{I}} = +f^2 P \frac{e^{-\lambda r}}{r} \left[(\boldsymbol{\sigma\rho}) \left(1 + \frac{1}{\lambda r} + \frac{1}{\lambda^2 r^2} \right) - \frac{(\boldsymbol{\sigma}\mathbf{r})\,(\boldsymbol{\rho}\mathbf{r})}{r^2} \left(1 + \frac{3}{\lambda r} + \frac{3}{\lambda^2 r^2} \right) \right]. \tag{37}$$

The total interaction operator is the sum of the two contributions

$$V = V_{\mathrm{I}} + V_{\mathrm{II}}. \tag{38}$$

In order to apply our result to the actual *problem of the proton-neutron interaction* we should have to insert V in the Schrödinger equation for the motion of these two particles. Equation (37) shows that there is a strong coupling between the orbital motion and the spins σ, ρ. This makes an exact solution of the Schrödinger equation very difficult. The operator V_{NP} leads only to a coupling between S and D-terms but not between S and P-terms. Since the range of the forces is so small that the D-term can be neglected it will be sufficient, for a rough physical discussion, to separate the orbital motion and the spin and to take the average value of the spin operators occurring in (37).

For the *wave function of the deuteron* we assume then a product of three functions

$$\Psi = \psi_l(1,2)\, T(1,2)\, U(1,2). \tag{39}$$

Here $\psi_l(1,2)$ depends only on the space coordinates of the two particles 1, 2. $T(1,2)$ is the spin function. $U(1,2)$ is the wave function describing which of the two particles, one and two, is a proton and which a neutron. Each of these three functions can be symmetrical or antisymmetrical in 1 and 2, but Ψ must be antisymmetrical. Thus there are four possibilities:

$$\left.\begin{array}{ll} \text{Triplet: (1)} & T\text{ sym.},\ \psi_l\text{ sym.},\ U\text{ antis.} \\ \text{(2)} & T\text{ sym.},\ \psi_l\text{ antis.},\ U\text{ sym.} \\ \text{Singlet: (3)} & T\text{ antis.},\ \psi_l\text{ sym.},\ U\text{ sym.} \\ \text{(4)} & T\text{ antis.},\ \psi_l\text{ antis.},\ U\text{ antis.} \end{array}\right\} \tag{40}$$

We confine ourselves to the consideration of S *terms*. Then ψ_l is *symmetrical* in the two particles, and we have only to consider the wave functions (1) and (3) of (40).

The operator P of equations (35) and (37) acts only on the wave function U and is

$$\left.\begin{array}{l} PU_{\text{sym.}} = +U_{\text{sym.}}, \\ PU_{\text{antis.}} = -U_{\text{antis.}} \end{array}\right\} \tag{41}$$

The operators $(\boldsymbol{\sigma\rho})$ and $(\boldsymbol{\sigma}\mathbf{r})(\boldsymbol{\rho}\mathbf{r})/r^2$ act on T. T has the following form:

$$\text{Singlet: } \alpha(1)\beta(2)-\alpha(2)\beta(1), \tag{42a}$$

$$\left.\begin{array}{l} \text{Triplet: } \alpha(1)\alpha(2), \\ \beta(1)\beta(2), \\ \alpha(1)\beta(2)+\alpha(2)\beta(1), \end{array}\right\} \tag{42b}$$

α, β are the two spin functions for each particle. Since the S-term is spherically symmetrical the average of the spin operators over all directions results, for the ^{3}S-term, in taking the average over all three spin functions (42*b*). We then obtain

$$\left.\begin{array}{l} \text{Singlet: } (\boldsymbol{\sigma\rho})T = -3T, \\ (\boldsymbol{\sigma}\mathbf{r})(\boldsymbol{\rho}\mathbf{r})/r^2T = -T, \\ \text{Triplet: } (\boldsymbol{\sigma\rho})T = +T, \\ (\boldsymbol{\sigma}\mathbf{r})(\boldsymbol{\rho}\mathbf{r})/r^2T = \tfrac{1}{3}T. \end{array}\right\} \tag{43}$$

Inserting now (41) and (43) into the expression for V_{NP} (35) and (37) taking into account the symmetry properties of the wave functions (40) we obtain for the mean value of V_{NP}:

$$\left.\begin{array}{l} {}^3\text{S}: V_{NP} = -\dfrac{e^{-\lambda r}}{r}(g^2+\tfrac{2}{3}f^2), \\ {}^1\text{S}: V_{NP} = -\dfrac{e^{-\lambda r}}{r}(2f^2-g^2). \end{array}\right\} \tag{44}$$

Thus we find that the *potential for the* ^{3}S-*term is always attractive.* For the ^{1}S term V is attractive if $2f^2 > g^2$. The experiments suggest that the proton-neutron potential in the lowest state ^{3}S can be represented by a hole of width 3×10^{-13} cm. and a depth of 25×10^6 e-volts. For the ^{1}S-state the potential is approximately half of that of the ^{3}S-state. Hence we obtain at once

$$f \doteqdot g. \tag{45}$$

The range of the nuclear forces is $1/\lambda = m_0 c/\hbar$. The mass of the heavy electron will therefore be

$$m_0 \doteqdot 100 \text{ electron masses.} \tag{46}$$

As "depth" of the potential hole we can define the quantity $(g^2 + \frac{2}{3}f^2)\lambda$. Taking into account equation (45) we find for g and f (or for the dimensionless quantity $g^2/\hbar c$)

$$g \doteqdot f \doteqdot 5 \text{ electron charges,} \quad \frac{g^2}{\hbar c} \doteqdot \frac{1}{6}. \tag{47}$$

All these figures are of course only very rough.

We have found that our scheme makes it possible to account in a reasonable way for the nuclear forces, including their right spin dependence. Although in our theory three constants are available, this success is by no means trivial. In the scalar theory (Yukawa and Sakata 1937) it turns out, for instance, that the ^{3}S-state is always repulsive and the ^{1}S-state attractive with the same absolute value, which is contrary to the experiments. Further the scalar theory does not allow for the magnetic moments of the proton and neutron. Only the vector scheme leads to a qualitative agreement with the experiments.

3. THE PROTON-PROTON INTERACTION

From scattering experiments (Tuve, Heydenburg and Hafstadt 1936; Breit, Condon and Present 1936) it is known that there are also short range forces between two protons which seem to be of the same order of magnitude as the neutron-proton forces. Although the experiments do not allow any conclusions as to the exact form and magnitude of these forces it has been suggested by Breit, Condon and Present that the forces between all nuclear particles are exactly identical.

In our theory we have so far only assumed an interaction by which *charged* particles are emitted by a proton or neutron with the corresponding change of charge of the latter particles. It is obvious that this interaction leads to a proton-proton force only in the fourth order of approximation. Although we shall see that the latter is not at all small, it is, of course, not to be expected that it is exactly the same as the neutron-proton force. It would, however, be possible to obtain the same forces irrespective of the charge of the nuclear particles if we would introduce a *neutral* particle of the same mass as that of the heavy electron, with a similar interaction with the nuclear particles to that given by equations (30).

It is quite possible that those particles exist. The cosmic-ray experiments, however, do not give any information yet about the existence of these particles, neither in the affirmative nor in the negative sense. Since the exact equality of the nuclear forces is by no means established experimentally, it seems in any case desirable to work out the consequences of our theory for the proton-proton forces in the fourth order of approximation without introducing neutral particles.

In the preceding section we have seen that the constant responsible for the nuclear forces, g^2 (or f^2) is of the order of $25e^2$. The second, fourth ... etc., orders of approximation represent an expansion in a power series of $g^2/\hbar c = 1/6$ which is by no means small compared with unity. Thus the proton-proton forces will at least be comparable with the proton-neutron forces.

The perturbation energy is in the fourth approximation given by the formula

$$V_{PP} = -\sum \frac{H'_{A\,\mathrm{I}} H'_{\mathrm{I}\,\mathrm{II}} H'_{\mathrm{II}\,\mathrm{III}} H'_{\mathrm{III}\,A}}{E_{\mathrm{I}} E_{\mathrm{II}} E_{\mathrm{III}}}. \tag{48}$$

Here I, II, III are intermediate states in which one or two heavy electrons are emitted and H' is the corresponding matrix element of (30). The intermediate states are, for instance, the following:

(i) Proton 1 has emitted a positive heavy electron with wave number $\mathbf{k}_1$.

$$E_{\mathrm{I}}^2 = \epsilon_1^2 = \hbar^2 c^2(k_1^2 + \lambda^2).$$

(ii) Proton 2 has emitted a positive heavy electron with wave number $\mathbf{k}_2$.

$$E_{\mathrm{II}} = \epsilon_1 + \epsilon_2. \quad \epsilon_2^2 = \hbar^2 c^2(k_2^2 + \lambda^2).$$

(iii) The positive electron $\mathbf{k}_1$ is absorbed by proton 2.

$$E_{\mathrm{III}} = \epsilon_2.$$

The order of emission and absorption can be permuted for each $\mathbf{k}_1$, $\mathbf{k}_2$, and we obtain in this way four different sets of intermediate states. The matrix elements do not depend on this order.

There is another set of intermediate states in which a proton emits subsequently a positive and a negative heavy electron which are then absorbed by the second proton. The denominator in (48) is for this intermediate state

$$\epsilon_+^2(\epsilon_+ + \epsilon_-).$$

This intermediate state occurs twice since the roles of the two protons can be interchanged. The matrix elements are the same as for the first group of intermediate states. Writing ϵ_1, ϵ_2 (or ϵ_2, ϵ_1) instead of ϵ_+, ϵ_- we obtain for the sum over all these intermediate states

$$\frac{2}{\epsilon_1\epsilon_2(\epsilon_1+\epsilon_2)}+\frac{1}{\epsilon_1^2(\epsilon_1+\epsilon_2)}+\frac{1}{\epsilon_2^2(\epsilon_1+\epsilon_2)}+\frac{2}{\epsilon_+^2(\epsilon_+ +\epsilon_-)}=\frac{2\epsilon_1\epsilon_2}{\epsilon_1^2-\epsilon_2^2}\left(\frac{1}{\epsilon_2^3}-\frac{1}{\epsilon_1^3}\right).$$

Since the expression (48) is symmetrical in ϵ_1, ϵ_2 we obtain for V_{PP}

$$V_{PP} = -4\Sigma\epsilon_1\epsilon_2\frac{H'_{A\,\mathrm{I}}H'_{\mathrm{I}\,\mathrm{II}}H'_{\mathrm{II}\,\mathrm{III}}H'_{\mathrm{III}\,A}}{(\epsilon_1^2-\epsilon_2^2)\epsilon_2^3}, \tag{49}$$

where the sum has now to be extended over all directions and values of the two wave vectors $\mathbf{k}_1$ and $\mathbf{k}_2$, and all directions of polarizations. The matrix elements $H'_{A\,\mathrm{I}}$, etc., have to be inserted from equation (32). For $H'_{A\,\mathrm{I}}H'_{\mathrm{II}\,\mathrm{III}}$ and for $H'_{\mathrm{I}\,\mathrm{II}}H'_{\mathrm{III}\,A}$ we obtain very similar expressions as those obtained for the neutron-proton force in the preceding section. After carrying out the summation over all directions of polarization we obtain

$$H'_{A\,\mathrm{I}}H'_{\mathrm{II}\,\mathrm{III}} = \frac{2\pi\hbar^2c^2}{\epsilon_1\lambda^2}\{g^2k_1^2+f^2[k_1^2(\boldsymbol{\sigma}\boldsymbol{\rho})-(\boldsymbol{\sigma}\mathbf{k}_1)(\boldsymbol{\rho}\mathbf{k}_1)]\}e^{i(\mathbf{k}_1\mathbf{r})}. \tag{50}$$

Here $\mathbf{r} = \mathbf{r}_1-\mathbf{r}_2$ is the radius vector of the distance between the two protons. $H'_{\mathrm{I}\,\mathrm{II}}H'_{\mathrm{III}\,A}$ is obtained by replacing $\mathbf{k}_1$ by $\mathbf{k}_2$ and $\mathbf{r}$ by $-\mathbf{r}$.

In all physical problems it is sufficient to consider the interaction of two protons only in the S-state because the range of the forces is very small. Our further calculations will be very much simplified if we confine ourselves to the evaluation of V_{PP} for the ^{1}S-state. In this case the spin operators occurring in (50) are

$$(\boldsymbol{\sigma}\boldsymbol{\rho}) = -3, \quad (\boldsymbol{\sigma}\mathbf{k}_1)(\boldsymbol{\rho}\mathbf{k}_1) = -\mathbf{k}_1^2.$$

For V_{PP} we thus obtain

$$V_{PP} = -4\left(\frac{2\pi\hbar^2c^2}{\lambda^2}\right)^2(g^2-2f^2)^2\Sigma\frac{k_1^2k_2^2e^{i(\mathbf{k}_1\mathbf{r})}e^{-i(\mathbf{k}_2\mathbf{r})}}{(\epsilon_1^2-\epsilon_2^2)\epsilon_2^3}.$$

Replacing the sum by an integral and carrying out the integration over the angles the sum becomes

$$\Sigma = \frac{1}{4\pi^4(\hbar c)^5 r^2}\frac{1}{}\int_0^\infty\int_0^\infty \frac{k_1^3 k_2^3 \sin k_1 r \sin k_2 r \, dk_1 \, dk_2}{(k_1^2 - k_2^2)(k_2^2+\lambda^2)^{\frac{3}{2}}}.$$

If we put $k_1 r = x$, $k_2 r = y$ the integration over x can easily be performed and yields for V_{PP}:

$$V_{PP} = -\frac{2}{\pi\hbar c}\frac{(g^2-2f^2)^2}{\lambda^4 r^5}\int_0^\infty \frac{y^5 \sin y \cos y \, dy}{(y^2+\lambda^2 r^2)^{\frac{3}{2}}}. \tag{51}$$

The integral in (51) leads to Hankel functions. It can be reduced to the integral

$$\int_0^\infty \frac{x \sin \alpha x \, dx}{(x^2+\beta^2)^{\frac{3}{2}}} = i\frac{\pi}{2}\alpha H_0^{(1)}(i\beta\alpha)$$

and its derivatives. Using $2\sin y \cos y = \sin 2y$ and the well known relations between Hankel-functions we obtain finally

$$V_{PP} = -\frac{(g^2-2f^2)^2}{\hbar c r}\left[\left(1-\frac{1}{(2\lambda r)^2}\right)iH_0^{(1)}(2i\lambda r) + \left(1+\frac{1}{(2\lambda r)^2}\right)\frac{H_1^{(1)}(2i\lambda r)}{\lambda r}\right]. \tag{52}$$

We compare this formula with the neutron-proton force (equation (44)) in the ^{1}S-state and find, since $f \doteqdot g$

$$V_{PP} = V_{NP}\frac{g^2}{\hbar c}e^{\lambda r} \times [\text{Hankel functions}]. \tag{53}$$

$g^2/\hbar c$ is of the order 1/6. To discuss the factor containing the Hankel functions we give the values of this factor for a few values of λr:†

λr	$e^{\lambda r} \times$ [Hankel functions]
0·1	−900
0·2	− 70
0·5	− 2·5
1	− 0·2

Since V_{NP} is attractive we find that the *proton-proton force* is *always repulsive*. For very small distances the absolute value of the force is very much larger than that of the proton-neutron force and decreases very rapidly with r. Thus the protons behave nearly like two rigid spheres with a diameter d of about

$$d = 1/2\lambda. \tag{54}$$

The fact that the fourth order of approximation is greater than the second order for small distances means, of course, that the *whole theory*

---

† Cf. Jahnke-Emde (1933, pp. 286, 287).

diverges for small distances and our results can therefore only have a very qualitative significance. As a main result of our theory we can only say that the nuclear particles will have a finite radius of the order of equation (54). In this connexion we want to emphasize that the fourth order of the neutron-proton forces does not lead to any exchange force. The only way of exchanging two heavy electrons between a proton and a neutron is the successive emission of a positive and negative heavy electron by the proton or neutron and re-absorption by the other particle. It does obviously not lead to any exchange of charge. It would lead to a repulsion at very small distances of about half of the proton-proton repulsion.

The proton-proton force, as calculated in this section, leads to a strong scattering of protons by protons. It is not possible to say without detailed calculations whether the scattering due to these forces is of the order of magnitude observed experimentally. The authors mentioned above, however, concluded from the experimental data that the proton-proton force is mainly *attractive*. If this result is established it would inevitably lead to the conclusion, that we have also to introduce *neutral particles* (*"Neutretto's"*) of the same mass as and similar properties to those of the heavy electrons. The theory of these particles can be developed on the same lines as for the charged particles. We shall not, however, discuss this question in detail here. We want to emphasize that the forces calculated in this section will exist in any case. The proton-proton force can therefore only be attractive for distances larger than $1/2\lambda$ and will in any case be repulsive for smaller distances.

We may also mention that for a nucleus composed of three or more particles, forces also exist corresponding to a "triple exchange" of a heavy electron between three of the nuclear particles. These forces are of the same order as the fourth order proton-proton forces considered above.

4. THE MAGNETIC MOMENT

In this section we show that the virtual emission of a heavy electron leads to an additional magnetic moment of the proton and the neutron. To calculate this additional moment we proceed in the following way: We consider a proton placed in a weak homogeneous magnetic field H with spin momentum $+\frac{1}{2}$ in the direction of the magnetic field. The virtual emission and reabsorption of the heavy electron gives rise to a certain self-energy W depending on the magnetic field H. Expanding W in a power series of H

$$W = W_0 + \mu' H + \dots, \tag{55}$$

the factor of H gives the additional moment of the proton. Besides this surplus moment we shall assume the ordinary moment of one Bohr nuclear magneton μ_0 for the proton.

It follows immediately from the interaction (30) that only the term H'_{I} can give a contribution to the magnetic moment. H'_{II} does not contain the spin $\boldsymbol{\sigma}$ and cannot, therefore, give rise to a magnetic moment connected with the direction of the spin. In fact it would turn out that the term proportional to H in (55) is precisely equal to zero.

The magnetic moment, therefore, is due to the emission and reabsorption of *transverse waves*.

Since we are interested in the magnetic energy of the emitted heavy electron, it will be convenient to introduce polar co-ordinates r, θ, ϕ, and to expand the transverse waves into spherical waves with the polar axis in the direction of the field (z-axis). These waves can be classified according to their angular momentum m about the z-axis. Only waves with $m \neq 0$ will have a magnetic energy and therefore contribute to the magnetic moment. This expansion problem is almost identical with the expansion of electromagnetic waves into spherical waves in Maxwell's theory. The only difference is the fact that the frequency is now connected with the wave number by equation (11) whereas in Maxwell's theory $\nu = kc$. The latter problem has been treated already (Heitler 1936*b*). Our ϕ_i corresponds to the vector potential **A** and curl ϕ_i to the magnetic field strength curl **A**. For our interaction H'_{I} we need curl ϕ at the position of the proton ($r = 0$) which therefore corresponds to the magnetic field strength for $r = 0$.

The spherical electromagnetic waves can be divided into two groups representing an electrical and magnetic multipole radiation. Each of them is characterized by two integral "quantum numbers" l and m, where

$$-l \leqslant m \leqslant l, \quad l = 0, 1, 2, \ldots$$

m is proportional to the angular momentum about the z-axis. The expressions for the magnetic field strength for all these cases are explicitly given in the paper mentioned above. It turns out, however, that for $m \neq 0$ the magnetic field strength of the electrical multipole waves vanishes for $r = 0$. For the magnetic multipole wave the magnetic field strength is given by the formulae (28) and (29) of the paper quoted above:

$$\operatorname{curl}_x\phi \pm i\operatorname{curl}_y\phi = \pm ike^{i\nu t}q\left\{\left[\frac{l(l\pm m+2)(l\pm m+1)}{(l+1)(2l+1)(2l+3)}\right]^{\frac{1}{2}} Y_{l+1}^{m\pm1}f_{l+1} - \left[\frac{(l+1)(l\mp m-1)(l\mp m)}{l(2l+1)(2l-1)}\right]^{\frac{1}{2}} Y_{l-1}^{m\pm1}f_{l-1}\right\}, \quad (56)$$

where Y_l^m are the normalized spherical harmonics, and

$$f_l = C\frac{J_{l+\frac{1}{2}}(kr)}{\sqrt{(kr)}}$$

$J_{l+\frac{1}{2}}$ are Bessel functions, C is a constant which has to be taken in such a way that f_l is a normalized wave function. If this function is normalized for a sphere with radius 1,† then

$$C = k\sqrt{(8\pi c^2)}.$$

In formula (56) q is the quantized amplitude of the oscillator representing the heavy electron (cf. equation (25)). We have written down only the x- and y-components. The z-components consist of terms proportional to $Y_{l+1}^m f_{l+1}$ and $Y_{l-1}^m f_{l-1}$ and will be seen to give no contribution to the magnetic moment.

For $r = 0$ all $f_l(0)$ vanish except $f_0(0)$. Thus it follows from (56) that only the wave with $l = 1$, $|m| = 1$ gives a contribution. We then obtain for $r = 0$:

$$\left.\begin{aligned} m = +1:\ & \operatorname{curl}_x\phi + i\operatorname{curl}_y\phi = 0, \\ & \operatorname{curl}_x\phi - i\operatorname{curl}_y\phi = i\frac{2}{\sqrt{3}}kqY_0^0 f_0(0), \\ m = -1:\ & \operatorname{curl}_x\phi + i\operatorname{curl}_y\phi = -i\frac{2}{\sqrt{3}}kqY_0^0 f_0(0), \\ & \operatorname{curl}_x\phi - i\operatorname{curl}_y\phi = 0. \end{aligned}\right\} \tag{57}$$

The interaction H_1' can be written in the following form:

$$(\boldsymbol{\sigma}\operatorname{curl}\phi) = \sigma_z\operatorname{curl}_z\phi + \tfrac{1}{2}(\sigma_x + i\sigma_y)(\operatorname{curl}_x - i\operatorname{curl}_y)\phi + \tfrac{1}{2}(\sigma_x - i\sigma_y)(\operatorname{curl}_x + i\operatorname{curl}_y)\phi.$$

According to (31) we have

$$\left.\begin{aligned} \tfrac{1}{2}(\sigma_x + i\sigma_y) &= \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \\ \tfrac{1}{2}(\sigma_x - i\sigma_y) &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \end{aligned}\right\} \tag{58}$$

Thus it follows, together with (25), (29), (57), that a proton with spin $+\frac{1}{2}$ can emit only a positive electron with angular momentum $+1$, transforming itself into a neutron with spin $-\frac{1}{2}$ and vice versa. There are, of course, also transitions in which a heavy electron with angular momentum $m = 0$ is emitted but, as mentioned above, we are not interested in them here.

---

† As in equation (8) we normalize $Y_l^m f_l$ to $4\pi c^3$.

The self energy is given by the well-known formula

$$W = -\Sigma \frac{H'_{0n} H'_{n0}}{E_n - E_0}, \tag{59}$$

where E_n is the energy of the emitted heavy electron in the magnetic field and E_0 is the energy in the initial state, i.e. the energy of the Dirac proton in the magnetic field. For the present we are interested in the contribution of the heavy electron to the magnetic energy and shall neglect E_0. We shall consider the effect of the Bohr magneton of the proton below.

The energy E_n is given by†

$$E_n^2 = \hbar^2 c^2 \left[(k^2 + \lambda^2) - \frac{eHm}{\hbar c} \right], \quad m = 1 \tag{60}$$

neglecting terms in H^2. We insert in (59) H'_{0n} from equations (30a) and (57) and q from (25). In H'_{n0} the matrices $\sigma_x + i\sigma_y$, etc., have to be replaced by their hermitian conjugates. Furthermore the number of states in the interval dk of the wave number k is dk/π. Replacing the sum by an integral and inserting $Y_0^0 = 1/\sqrt{(4\pi)}$ and $f_0(0) = k\sqrt{(8\pi c^2)}$ we obtain for the self-energy

$$W = -\frac{4}{3\pi} \frac{f^2}{\lambda^2} \int \frac{k^4 dk}{k^2 + \lambda^2 - eH/\hbar c} = W_0 + \frac{4}{3\pi} \frac{f^2}{\lambda^2} \frac{eH}{\hbar c} \int_0^\infty \frac{k^4 dk}{(k^2 + \lambda^2)^2} \tag{61}$$

The factor of H represents the surplus moment of the proton. The sign is the same as for a positive charge with positive angular momentum. The integral diverges for large values of the wave number k. This shows that our theory cannot be valid for all wave-lengths. We have to introduce an upper limit k_0 for the wave number k marking the limit of validity of our theory. $1/k_0$ can hardly be other than of nuclear dimensions. We have already seen in our calculations of the nuclear forces (§ 3) that the higher orders of approximation diverge for distances $> 1/(2\lambda)$. We had to conclude that our theory, with the form (30) for the Hamiltonian, is restricted to distances $> 1/(2\lambda)$. Thus we think it reasonable to assume that the limit of validity of the theory for the magnetic moment is the same as for the nuclear forces and we shall therefore put

---

† It may be considered doubtful whether the magnetic energy is really given by $m\mu H$ (μ = relativistic magneton) as it is the case for the relativistic particle with no spin, described, for instance, by the scalar Klein-Gordon equation. Strictly speaking, in order to calculate the magnetic energy, we should require the field equations of the heavy electron in an external field. Our field equations can easily be generalized for this case.

A consistent canonical scheme has been proposed by Proca (1936). We have calculated the magnetic moment also by a consequent application of the Proca equations with Born's approximation, and have obtained the same result as in the text, apart from a factor 2 which arises from heavy electron pairs created by the magnetic field (cf. Kemmer 1938). This factor 2 does not play any role since the absolute value of the magnetic moment depends in any case on the quantity k_0 (see equation (63)) of which only the order of magnitude is defined. We have preferred to deduce the magnetic moment by the method used in the text because it is far more simple than the rigorous application of the Proca equations.

$$k_0 \text{ of the order } \lambda. \tag{62}$$

It will be seen that our theory leads then to the right order of magnitude of the surplus magnetic moment of the proton.

Writing for the integral simply k_0 we obtain for the surplus moment (dividing by the Bohr nuclear moment)

$$\frac{\mu'}{\mu_0} = \frac{8}{3\pi} \frac{f^2}{\lambda^2} \frac{Mk_0}{\hbar^2} = \frac{8}{3\pi} \frac{f^2}{\hbar c} \frac{M}{m_0} \frac{k_0}{\lambda}. \tag{63}$$

If we insert for f *the value obtained from the nuclear forces* (§ 2) $f^2/\hbar c = 1/6$ and for m_0 150 electron masses we obtain $\mu'/\mu_0 \fallingdotseq 2k_0/\lambda$. *This is of the right order of magnitude* if $k_0 \sim \lambda$ as we have suggested above.

The total magnetic moment of the proton will be

$$\mu_P = \mu' + \mu_0.$$

For the neutron the calculation is entirely analogous. The sign is the opposite because a negative electron is emitted. Therefore

$$\mu_N = -\mu'.$$

It would follow from these considerations that $\mu_P - |\mu_N| = \mu_0$. This is approximately the case since the magnetic moment of the deuteron is $\mu_D = \mu_P - |\mu_N| = 0{\cdot}85\mu_0$. The departure of this figure from $1 \times \mu_0$ can be explained by taking into account the magnetic energy of the Dirac proton $E_0 = \mu_0 H$ in equation (59) which we have so far neglected. For the neutron a similar term has to be taken into account since we have in the intermediate state a proton with spin direction $-\frac{1}{2}$ leading to a magnetic energy

$$-E_0 = -\mu_0 H.$$

Similar terms would even occur in transitions to electronic states with $m = 0$ since also in this case† the proton with magnetic energy $+\mu_0 H$ is transformed into a neutron with no magnetic energy (correspondingly a momentum. The interaction with the nuclear particles is found by relativistic arguments and contains—apart from the mass m_0—two arbitrary constants g and f, both with dimensions of an electric charge.

---

† Transitions of this kind will also be caused by the interaction with longitudinal waves, equation (30*b*).

With this scheme we have calculated:

(1) The neutron-proton force. It is an exchange force and has a range $1/\lambda = \hbar/m_0c$. In the ^{3}S-state the force is always attractive. (This would not be the case for a scalar wave function.) g and f can be chosen so that the ^{3}S and ^{1}S-states have the right position. The experiments suggest $g \risingdotseq f \risingdotseq 5$ electron charges.

(2) The magnetic moments μ_P and μ_N of the proton and neutron. They are found to be of nuclear dimensions and have the right sign.

(3) The mass m_0 can be determined independently from the range of the neutron-proton force and from the magnetic moments. In both ways we find $m_0 \risingdotseq 100$ electron masses.

(4) The proton-proton force is obtained only in the fourth order of approximation and leads to a strong repulsion for distances less than $1/(2\lambda)$. Attraction and equality with the neutron-proton force could be attained by introducing also neutral particles with mass m_0.

(5) The theory leads to a diverging self-energy of the proton and neutron.

REFERENCES

Bhabha 1938 *Nature, Lond.*, **141**, 117.
Breit, Condon and Present 1936 *Phys. Rev.* **50**, 825.
Fröhlich and Heitler 1938 *Nature, Lond.*, **141**, 37.
Heitler 1936*a* "Quantum Theory of Radiation." Oxford.
— 1936*b* *Proc. Camb. Phil. Soc.* **32**, 112.
Iwanenko 1934 *Nature, Lond.*, **133**, 981.
Jahnke-Emde 1933 "Tables of Functions." Leipzig.
Kemmer 1938 *Proc. Roy. Soc.* A, **166**, 127.
Neddermeyer and Anderson 1937 *Phys. Rev.* **51**, 884.
Pauli and Weisskopf 1934 *Helv. phys. Acta*, **7**, 709.
Proca 1936 *J. Phys. Radium*, **7**, 347.
Tamm 1934 *Nature, Lond.*, **133**, 981.
Tuve, Heydenburg and Hafstadt 1936 *Phys. Rev.* **51**, 1023.
Wick 1935 *R.C. Accad. Lincei*, **21**, 170.
Yukawa 1935 *Proc. Phys. Math. Soc. Japan*, **17**, 48.
Yukawa and Sakata 1937 *Proc. Phys. Math. Soc. Japan*, **19**, 1084.

The lift and moment on a flat plate in a stream of finite width

By T. H. Havelock, F.R.S.

(*Received* 8 *February* 1938)

1. The problem of the lift on a flat plate in a stream between parallel rigid walls has been solved in an exact form, by using a suitable conformal transformation, by Tomotika (1934), who also gives an expansion for the lift in the particular case when the mid-point of the plate is midway between the walls; a similar solution for the moment on the plate does not seem to have been given. The method used in the following paper is quite different and is, perhaps, of sufficient interest to justify further examination of the problem. The flat plate is treated as the limiting case of an elliptic cylinder, and the method of solution leads directly to expansions for the lift and for the moment suitable for any position of the plate subject to the parameters being within the range necessary for convergence. Moreover, by a simple modification, expansions for lift and moment are obtained when the stream is bounded by parallel free surfaces, taking the boundary condition in an approximate form; and a further modification gives the corresponding results when one surface is rigid and the other free. A brief examination is also made of the moment for an elliptic cylinder.

GENERAL EXPRESSIONS

2. Consider the two-dimensional motion due to a cylinder placed in a uniform stream bounded by plane parallel walls, including circulation round the cylinder. Let C be the contour of the cross-section of the cylinder, and take the origin O so that the parallel walls are given by $y = a$, and $y = -b$, respectively. To simplify the argument, we assume that a position can be found for O such that a circle can be drawn, with centre O, entirely in the liquid and enclosing the contour C.

With w for the complex potential function, we take

$$\frac{dw}{dz} = c + \sum_0 \frac{i^{n+1} n!\, A_n}{z^{n+1}} + \frac{dw_1}{dz}. \tag{1}$$

In (1), c is the velocity of the stream, in the negative direction of Ox, the series is a suitable expansion for the singularities of the potential function within the contour C, and the last term is to be determined so as to satisfy the boundary conditions on the walls. These conditions are

$$I\frac{dw}{dz} = 0; \quad z = x+ia,\ z = x-ib, \tag{2}$$

where I denotes the imaginary part.

To satisfy these conditions we replace the series in (1) by

$$\int_0^\infty F(\kappa)\,e^{i\kappa z}\,d\kappa, \quad \text{for } z = x+ia,$$

$$-\int_0^\infty F(-\kappa)\,e^{-i\kappa z}\,d\kappa, \quad \text{for } z = x-ib,$$

where $$F(\kappa) = \sum_0 A_n\kappa^n. \tag{3}$$

We may build up an expression for dw_1/dz by successive images. Taking the expressions in (3), a single reflexion at a plane wall changes $F(\kappa)$ into the conjugate complex $F^*(\kappa)$; if the reflexion is at the upper wall ($y = a$) the contribution to dw_1/dz valid in the liquid is

$$\int_0^\infty F^*(\kappa)\,e^{-i\kappa z-2\kappa a}\,d\kappa, \tag{4}$$

while if the reflexion is at the lower wall ($y = -b$), the corresponding form is

$$-\int_0^\infty F^*(-\kappa)\,e^{i\kappa z-2\kappa b}\,d\kappa. \tag{5}$$

Taking successive reflexions at the two walls, the contributions of the infinite sets of image systems may be summed, and we obtain finally

$$\begin{aligned}\frac{dw}{dz} = c + \sum_0 \frac{i^{n+1}n!\,A_n}{z^{n+1}} &+ \int_0^\infty \frac{F^*(\kappa)\,e^{-i\kappa z-2\kappa a}-F^*(-\kappa)\,e^{i\kappa z-2\kappa b}}{1-e^{-2\kappa d}}\,d\kappa \\ &- \int_0^\infty \frac{F(-\kappa)\,e^{-i\kappa z-2\kappa d}-F(\kappa)\,e^{i\kappa z-2\kappa d}}{1-e^{-2\kappa d}}\,d\kappa,\end{aligned} \tag{6}$$

where $d = a+b$, and $F(\kappa) = \Sigma A_n\kappa^n$.

It may easily be verified directly, by using (4) and (5), that (6) satisfies the boundary conditions (2).

3. We now calculate the forces on the cylinder from the expression

$$X - iY = \tfrac{1}{2}\rho i \oint \left(\frac{dw}{dz}\right)^2 dz. \tag{7}$$

We write (6) in the form

$$\frac{dw}{dz} = c + \Sigma \frac{i^{n+1} n!\, A_n}{z^{n+1}} + \Sigma B_n z^n, \tag{8}$$

where
$$B_n = \int_0^\infty \frac{(-i)^n F^*(\kappa)\, e^{-2\kappa a} - i^n F^*(-\kappa)\, e^{-2\kappa b}}{1 - e^{-2\kappa d}} \frac{\kappa^n}{n!} d\kappa$$
$$- \int_0^\infty \frac{\{(-i)^n F(-\kappa) - i^n F(\kappa)\}\, e^{-2\kappa d}}{1 - e^{-2\kappa d}} \frac{\kappa^n}{n!} d\kappa. \tag{9}$$

From (7) and (8), we obtain

$$X - iY = -2\pi\rho(icA_0 + \Sigma i^{n+1} n!\, A_n B_n). \tag{10}$$

If Γ is the circulation round the cylinder, we have $\Gamma = 2\pi A_0$; further, using (9), we easily obtain $X = 0$, and

$$Y = \rho c\Gamma + 2\pi\rho \int_0^\infty \frac{F(\kappa)\, F^*(\kappa)\, e^{-2\kappa a} - F(-\kappa)\, F^*(-\kappa)\, e^{-2\kappa b}}{1 - e^{-2\kappa d}} d\kappa. \tag{11}$$

For the moment about the origin we have

$$M = -\tfrac{1}{2}\rho R \oint z \left(\frac{dw}{dz}\right)^2 dz \tag{12}$$

$$= 2\pi\rho R i\{cA_1 + \Sigma i^n (n+1)!\, A_{n+1} B_n\}, \tag{13}$$

where R denotes the real part.

Using (9), this may be expressed in the form

$$M = 2\pi\rho R i\left[cA_1 + \int_0^\infty \frac{F'(\kappa)\, F^*(\kappa)\, e^{-2\kappa a} - F'(-\kappa)\, F^*(-\kappa)\, e^{-2\kappa b}}{1 - e^{-2\kappa d}} d\kappa \right.$$
$$\left. - \int_0^\infty \frac{\{F'(\kappa)\, F(-\kappa) - F'(-\kappa)\, F(\kappa)\}\, e^{-2\kappa d}}{1 - e^{-2\kappa d}} d\kappa \right]. \tag{14}$$

To complete the solution of the problem in any given case we have to determine the function $F(\kappa)$ so that the boundary condition of zero normal velocity is satisfied over the contour C.

ELLIPTIC CYLINDER

4. We take the contour C to be an ellipse, of semi-axes a' and b', with its major axis making an acute angle θ with the positive direction of Ox, and we take the origin at the centre of the ellipse.

In terms of a complex variable $\zeta\ (=\xi+i\eta)$, we take

$$z = pe^{i\theta}\cosh\zeta, \tag{15}$$

and the contour C is given by

$$\xi=\xi_0;\ p\cosh\xi_0 = a';\ p\sinh\xi_0 = b'. \tag{16}$$

We now write

$$\frac{dw}{d\zeta} = cpe^{i\theta}\sinh\zeta+\frac{i\Gamma}{2\pi}-\sum_1 i^n b_n e^{-n\zeta}+pe^{i\theta}\sinh\zeta\frac{dw_1}{dz}, \tag{17}$$

the second and third terms being in a suitable form in the elliptic coordinates; to obtain $F(\kappa)$ in terms of the new coefficients b_n we have to compare these two terms with the series in (6), noting that

$$dz/d\zeta = pe^{i\theta}\sinh\zeta.$$

For this purpose we put the series in (6) into the form in (3) valid for the upper surface; under the same condition it can be shown that

$$\frac{i\Gamma}{2\pi}-\Sigma i^n b_n e^{-n\zeta} = pe^{i\theta}\sinh\zeta\int_0^\infty\left\{\frac{\Gamma}{2\pi}J_0(\kappa pe^{i\theta})+i\Sigma b_n J_n(\kappa pe^{i\theta})\right\}e^{i\kappa z}\,d\kappa. \tag{18}$$

Hence, by comparison, we obtain

$$F(\kappa)=\frac{\Gamma}{2\pi}J_0(\kappa pe^{i\theta})+i\Sigma b_n J_n(\kappa pe^{i\theta}), \tag{19}$$

and
$$\begin{aligned}\frac{dw}{d\zeta} &= cpe^{i\theta}\sinh\zeta+\frac{i\Gamma}{2\pi}-\Sigma i^n b_n e^{-n\zeta}\\ &\quad+pe^{i\theta}\sinh\zeta\int_0^\infty\frac{F^*(\kappa)\,e^{-i\kappa z-2\kappa a}-F^*(-\kappa)\,e^{i\kappa z-2\kappa b}}{1-e^{-2\kappa d}}\,d\kappa\\ &\quad-pe^{i\theta}\sinh\zeta\int_0^\infty\frac{\{F(-\kappa)\,e^{-i\kappa z}-F(\kappa)\,e^{i\kappa z}\}\,e^{-2\kappa d}}{1-e^{-2\kappa d}}\,d\kappa.\end{aligned} \tag{20}$$

We now express this in the form

$$\frac{dw}{d\zeta}=\Sigma(C_n e^{n\zeta}+D_n e^{-n\zeta}), \tag{21}$$

by substituting in (20) the expansions

$$\kappa p e^{i\theta} e^{\pm i\kappa z} \sinh\zeta = 2\sum_1 (\pm i)^{n-1} nJ_n(\kappa p e^{i\theta}) \sinh n\zeta, \tag{22}$$

We obtain, for $n > 1$,

$$C_n = \int_0^\infty \frac{(-i)^{n-1} F^*(\kappa) e^{-2\kappa a} - i^{n-1} F^*(-\kappa) e^{-2\kappa b}}{1-e^{-2\kappa d}} nJ_n(\kappa p e^{i\theta}) \frac{d\kappa}{\kappa}$$

$$- \int_0^\infty \frac{\{(-i)^{n-1} F(-\kappa) - i^{n-1} F(\kappa)\} e^{-2\kappa d}}{1-e^{-2\kappa d}} nJ_n(\kappa p e^{i\theta}) \frac{d\kappa}{\kappa}, \tag{23}$$

$$D_n = -i^n b_n - \int_0^\infty \frac{(-i)^{n-1} F^*(\kappa) e^{-2\kappa a} - i^{n-1} F^*(-\kappa) e^{-2\kappa b}}{1-e^{-2\kappa d}} nJ_n(\kappa p e^{i\theta}) \frac{d\kappa}{\kappa}$$

$$+ \int_0^\infty \frac{\{(-i)^{n-1} F(-\kappa) - i^{n-1} F(\kappa)\} e^{-2\kappa d}}{1-e^{-2\kappa d}} nJ_n(\kappa p e^{i\theta}) \frac{d\kappa}{\kappa}, \tag{24}$$

while for $n = 1$, C_1 has the additional term $\frac{1}{2}cpe^{i\theta}$ and D_1 the additional term $-\frac{1}{2}cpe^{i\theta}$.

The boundary condition on the contour C is that the real part of $dw/d\zeta$ should be zero for $\xi = \xi_0$; this gives

$$D_n^* = -e^{2n\xi_0} C_n. \tag{25}$$

Using this in (23) and (24) we obtain an infinite set of equations for the coefficients b_n; these are, for $n > 1$,

$$ib_n = \int_0^\infty \frac{H(\kappa)}{1-e^{-2\kappa d}} \frac{d\kappa}{\kappa},$$

$$\begin{aligned} H(\kappa) = {} & \{nq^n F(\kappa) J_n(\kappa p e^{-i\theta}) + (-1)^n nF^*(\kappa) J_n(\kappa p e^{i\theta})\} e^{-2\kappa a} \\ & + \{(-1)^n nq^n F(-\kappa) J_n(\kappa p e^{-i\theta}) + nF^*(-\kappa) J_n(\kappa p e^{i\theta})\} e^{-2\kappa b} \\ & - n[\{q^n F^*(-\kappa) + (-1)^n q^n F^*(\kappa)\} J_n(\kappa p e^{-i\theta}) \\ & + \{(-1)^n F(-\kappa) + F(\kappa)\} J_n(\kappa p e^{i\theta})] e^{-2\kappa d}, \end{aligned} \tag{26}$$

with a similar expression for ib_1 including an additional term

$$-\tfrac{1}{2}cpe^{i\theta} + \tfrac{1}{2}cpqe^{-i\theta},$$

and with $q = e^{2\xi_0}$.

By using (19), these results may be combined into an integral equation for the function $F(\kappa)$; it is

$$F(\kappa) = \frac{\Gamma}{2\pi} J_0(\kappa p e^{i\theta}) - \tfrac{1}{2}cp(e^{i\theta} - qe^{-i\theta}) J_1(\kappa p e^{i\theta})$$

$$+ \int_0^\infty \frac{\{F(v) G_1 + F^*(v) G_2\} e^{-2va} + \{F^*(-v) G_3 + F(-v) G_4\} e^{-2vb}}{1-e^{-2vd}} \frac{dv}{v}$$

$$- \int_0^\infty \frac{\{F^*(-v) G_1 + F(-v) G_2 + F(v) G_3 + F^*(v) G_4\} e^{-2vd}}{1-e^{-2vd}} \frac{dv}{v}, \tag{27}$$

with
$$\left.\begin{aligned} G_1 &= \Sigma n q^n J_n(vpe^{-i\theta})\, J_n(\kappa p e^{i\theta}),\\ G_2 &= \Sigma(-1)^n\, nJ_n(vpe^{i\theta})\, J_n(\kappa pe^{i\theta}),\\ G_3 &= \Sigma nJ_n(vpe^{i\theta})\, J_n(\kappa pe^{i\theta}),\\ G_4 &= \Sigma(-1)^n\, nq^n J_n(vpe^{-i\theta})\, J_n(\kappa pe^{i\theta}).\end{aligned}\right\} \tag{28}$$

FLAT PLATE BETWEEN PARALLEL WALLS

5. We consider the limiting case obtained by making ξ_0 zero, that is, by putting $q = 1$ in the previous results. The cylinder reduces to a flat plate of width $2p$, at an angle θ to the direction of the stream, and with its mid-point at distances a, b from the upper and lower boundaries, respectively.

We write
$$\Gamma = 2\pi kcp \sin\theta; \quad F(\kappa) = cp\sin\theta f(\kappa). \tag{29}$$

The equation for $f(\kappa)$ is

$$\begin{aligned} f(\kappa) &= kJ_0(\kappa pe^{i\theta}) - iJ_1(\kappa pe^{i\theta})\\ &\quad + \int_0^\infty \frac{\{f(v)\,G_1 + f^*(v)\,G_2\}\,e^{-2va} + \{f^*(-v)\,G_3 + f(-v)\,G_4\}\,e^{-2vb}}{1-e^{-2vd}}\,\frac{dv}{v}\\ &\quad - \int_0^\infty \frac{\{f^*(-v)\,G_1 + f(-v)\,G_2 + f(v)\,G_3 + f^*(v)\,G_4\}\,e^{-2vd}}{1-e^{2vd}}\,\frac{dv}{v}, \end{aligned} \tag{30}$$

G_1, G_2, G_3, G_4 being given by (28) with $q = 1$.

We approximate to $f(\kappa)$ by successive substitution of approximations for $f(\kappa)$ in the integrals of (30), repeating the process as far as may be desired. Our object is to obtain the various quantities ultimately in power series in p/d, or alternatively in p/a or p/b, assuming these ratios to be less than unity. The expansion for $f(\kappa)$ is most readily obtained by replacing the Bessel-functions in (30) by their power series as far as necessary so as to give all terms up to a required order in the final results. We shall develop these expansions up to terms of order $(p/d)^4$; except for the length of the expressions, the expansions could readily be taken to a higher order. It is sufficient, for the present purpose, to take as the first approximation

$$f(v) = k - \tfrac{1}{2}ivpe^{i\theta} - \tfrac{1}{4}kv^2p^2e^{2i\theta} + \tfrac{1}{16}iv^3p^3e^{3i\theta} + \tfrac{1}{64}kv^4p^4e^{4i\theta}. \tag{31}$$

Further, to this order, it is sufficient to replace G_1 by

$$\begin{aligned} G_1 &= \tfrac{1}{2}\kappa pe^{i\theta}(\tfrac{1}{2}vpe^{-i\theta} - \tfrac{1}{16}v^3p^3e^{-3i\theta})\\ &\qquad + \tfrac{1}{8}\kappa^2p^2e^{2i\theta}(\tfrac{1}{4}v^2p^2e^{-2i\theta}) - \tfrac{1}{16}\kappa^3p^3e^{3i\theta}(\tfrac{1}{2}vpe^{-i\theta}), \end{aligned} \tag{32}$$

and G_2, G_3, G_4 by similar expressions.

We now obtain the result of putting in the integrals of (30) a typical term $v^n p^n e^{in\theta}$ or $iv^n p^n e^{in\theta}$ instead of $f(v)$; we then apply these results to each of the terms in (31) and repeat the process until we have obtained all the terms of the required order. For the integrations with respect to v which occur in the process we use the notation

$$r_n = \int_0^\infty \frac{e^{-2va} - (-1)^n e^{-2vb}}{1 - e^{2vd}} v^n dv, \tag{33}$$

$$r_n = \int_0^\infty \frac{e^{-2vd}}{1 - e^{-2vd}} v^n dv, \tag{34}$$

n not being zero in the second case; these integrals may be evaluated in finite form.

We now give the result of this process; we obtain

$$f(\kappa) = k - \tfrac{1}{2} i\kappa p B_1 e^{i\theta} - \tfrac{1}{4}\kappa^2 p^2 B_2 e^{2i\theta} + \tfrac{1}{16} i\kappa^3 p^3 B_3 e^{3i\theta} + \tfrac{1}{64}\kappa^4 p^4 B_4 e^{4i\theta} + \dots, \tag{35}$$

with

$$\left.\begin{aligned} B_1 &= 1 + kpr_0 \sin\theta + \tfrac{1}{2}p^2(r_1 - 2r_1' \cos 2\theta) \\ &\quad + \tfrac{1}{8}kp^3\{r_2(4\sin^3\theta - \sin\theta) + 4r_0(r_1 - 2r_1'\cos 2\theta)\sin\theta\} \\ &\quad - \tfrac{1}{8}p^4\{r_3 \cos 2\theta - 2r_3' \cos 4\theta - 2(r_1 - 2r_1' \cos 2\theta)^2\} + \dots, \\ B_2 &= k - \tfrac{1}{4}kp^2(r_1 - 2r_1') \cos 2\theta + \tfrac{1}{8}p^3 r_2 \sin\theta + \dots, \\ B_3 &= 1 + kpr_0 \sin\theta + \dots, \\ B_4 &= k + \dots. \end{aligned}\right\} \tag{36}$$

6. We consider in particular the case in which the circulation is such that the fluid velocity remains finite at the rear edge of the plate; the condition for this in an infinite stream is

$$\Gamma = 2\pi cp \sin\theta, \quad \text{or } k = 1.$$

Returning to the expression (21) for the elliptic cylinder, the condition requires that the imaginary part of $dw/d\zeta$ should be zero for $\zeta = \xi_0 + \pi i$. This gives, after putting $\xi_0 = 0$ for the flat plate,

$$\frac{i\Gamma}{2\pi} + \Sigma(-1)^n (C_n - C_n^*) = 0. \tag{37}$$

From (23) we obtain

$$\begin{aligned} \Sigma(-1)^n C_n &= -\tfrac{1}{2}cpe^{i\theta} \\ &\quad - \Sigma\int_0^\infty \frac{i^{n-1}F^*(\kappa)\, e^{-2\kappa a} - (-i)^{n-1} F^*(-\kappa)\, e^{-2\kappa b}}{1 - e^{-2\kappa d}}\, nJ_n(\kappa p e^{i\theta}) \frac{d\kappa}{\kappa} \\ &\quad + \Sigma\int_0^\infty \frac{\{i^{n-1}F(-\kappa) - (-i)^{n-1} F(\kappa)\}\, e^{-2\kappa d}}{1 - e^{-2\kappa d}}\, nJ_n(\kappa p e^{i\theta}) \frac{d\kappa}{\kappa}. \end{aligned} \tag{38}$$

Hence, writing

$$\left.\begin{aligned} A_1 &= J_0(\kappa pe^{i\theta})+iJ_1(\kappa pe^{i\theta}),\\ A_2 &= J_0(\kappa pe^{i\theta})-iJ_1(\kappa pe^{i\theta}),\end{aligned}\right\} \tag{39}$$

this gives

$$\Sigma(-1)^n C_n = -\tfrac{1}{2}cpe^{i\theta}-\tfrac{1}{2}pe^{i\theta}\int_0^\infty \frac{A_1F^*(\kappa)\,e^{-2\kappa a}-A_2F^*(-\kappa)\,e^{-2\kappa b}}{1-e^{-2\kappa d}}\,d\kappa$$
$$+\tfrac{1}{2}pe^{i\theta}\int_0^\infty \frac{\{A_1F(-\kappa)-A_2F(\kappa)\}\,e^{-2\kappa d}}{1-e^{-2\kappa d}}\,d\kappa. \tag{40}$$

Hence the equation for k, to give the required circulation $2\pi kcp\sin\theta$, is

$$k = 1+Ipe^{i\theta}\int_0^\infty \frac{A_1f^*(\kappa)\,e^{-2\kappa a}-A_2f^*(-\kappa)\,e^{-2\kappa b}}{1-e^{-2\kappa d}}\,d\kappa$$
$$-Ipe^{i\theta}\int_0^\infty \frac{\{A_1f(-\kappa)-A_2f(\kappa)\}\,e^{-2\kappa d}}{1-e^{-2\kappa d}}\,d\kappa, \tag{41}$$

with A_1, A_2 given by (39).

We substitute (35) in (41) and also use the power series for A_1 and A_2; carrying out the integrations and using the same notation as before, we find

$$\begin{aligned} k = 1 &+ kpr_0\sin\theta+\tfrac{1}{2}p^2r_1(k\cos 2\theta+B_1)-p^2r_1'(k+B_1)\cos 2\theta\\ &-\tfrac{1}{4}p^3r_2(k\sin 3\theta+B_1\sin\theta-B_2\sin\theta)\\ &-\tfrac{1}{16}p^4r_3(k\cos 4\theta+2B_1\cos 2\theta+2B_2+B_3\cos 2\theta)\\ &+\tfrac{1}{8}p^4r_3'(k+2B_1+2B_2+B_3)\cos 4\theta+\dots. \end{aligned} \tag{42}$$

Finally we substitute from (36) and solve the equation for k; we obtain

$$\left.\begin{aligned} k &= 1+a_1p+a_2p^2+a_3p^3+a_4p^4+\dots,\\ a_1 &= r_0\sin\theta,\\ a_2 &= r_0^2\sin^2\theta+r_1\cos^2\theta-2r_1'\cos 2\theta,\\ a_3 &= r_0^3\sin^3\theta+2r_0r_1\sin\theta\cos^2\theta-\tfrac{1}{4}r_2\sin 3\theta-4r_0r_1'\sin\theta\cos 2\theta,\\ a_4 &= r_0^4\sin^4\theta+3r_0^2r_1\sin^2\theta\cos^2\theta-\tfrac{1}{2}r_0r_2\sin\theta\sin 3\theta\\ &\quad+\tfrac{1}{4}r_1^2(3-6\sin^2\theta+4\sin^4\theta)-\tfrac{1}{16}r_3(2+3\cos 2\theta+\cos 4\theta)\\ &\quad-6r_0^2r_1'\sin^2\theta\cos 2\theta-3r_1r_1'\cos^2\theta\cos 2\theta+3r_1'^2\cos^2 2\theta+\tfrac{3}{4}r_3'\cos 4\theta. \end{aligned}\right\} \tag{43}$$

7. We may now obtain the lift from (11) and we express it in terms of the corresponding lift in an infinite stream; that is, we write

$$Y = L;\quad Y_0 = L_0 = 2\pi\rho c^2p\sin\theta.$$

Using (35) in (11) we obtain

$$L/L_0 = k + k^2 p r_0 \sin\theta + k p^2 r_1 B_1 \sin^2\theta - \tfrac{1}{4} p^3 r_2 (2kB_2 \cos 2\theta - B_1^2) \sin\theta$$
$$- \tfrac{1}{8} p^4 r_3 (kB_3 \sin 3\theta - 2B_1 B_2 \sin\theta) \sin\theta + \dots. \quad (44)$$

Substituting from (36) and (43) and collecting the terms we obtain

$$\left.\begin{aligned}
L/L_0 &= 1 + b_1 p + b_2 p^2 + b_3 p^3 + b_4 p^4 + \dots, \\
b_1 &= 2r_0 \sin\theta, \\
b_2 &= 3r_0^2 \sin^2\theta + r_1 - 2r_1' \cos 2\theta, \\
b_3 &= 4r_0^3 \sin^3\theta + 2r_0 r_1 (2\sin\theta - \sin^3\theta) - r_2 \sin\theta \cos 2\theta - 8r_0 r_1' \sin\theta \cos 2\theta, \\
b_4 &= 5r_0^4 \sin^4\theta + 3r_0^2 r_1 (3\sin^2\theta - 2\sin^4\theta) + \tfrac{3}{4} r_1^2 - \tfrac{3}{8} r_3 \cos 2\theta \\
&\quad - \tfrac{1}{2} r_0 r_2 (7\sin^2\theta - 12\sin^4\theta) - 18 r_0^2 r_1' \sin^2\theta \cos 2\theta - 3r_1 r_1' \cos 2\theta \\
&\quad + 3r_1'^2 \cos^2 2\theta + \tfrac{3}{4} r_3' \cos 4\theta.
\end{aligned}\right\} \quad (45)$$

The integrals given in (33) and (34) give for the coefficients,

$$\left.\begin{aligned}
r_0 &= \frac{\pi}{2d} \tan\alpha, \\
r_1 &= \left(\frac{\pi}{2d}\right)^2 \sec^2\alpha, \\
r_2 &= 2\left(\frac{\pi}{2d}\right)^3 \sec^2\alpha \tan\alpha, \\
r_3 &= 2\left(\frac{\pi}{2d}\right)^4 \sec^2\alpha (\sec^2\alpha + 2\tan^2\alpha), \\
r_1' &= \frac{\pi^2}{24d^2}; \quad r_3' = \frac{\pi^4}{240d^4},
\end{aligned}\right\} \quad (46)$$

with $\alpha = \pi(b-a)/2d$.

We may derive limiting cases from (45). If we make b and d infinite, we have a semi-infinite stream bounded by an upper plane wall; the limiting values of the coefficients are then

$$r_0 = \frac{1}{2a}; \quad r_1 = \frac{1}{4a^2}; \quad r_2 = \frac{1}{4a^3}; \quad r_3 = \frac{3}{8a^4}; \quad r_1' = r_3' = 0. \quad (47)$$

With these values we obtain

$$L/L_0 = 1+\frac{1}{2}\left(\frac{2p}{a}\right)\sin\theta+\frac{1}{16}\left(\frac{2p}{a}\right)^2(1+3\sin^2\theta)$$
$$+\frac{1}{32}\left(\frac{2p}{a}\right)^3(\sin\theta+3\sin^3\theta)-\frac{1}{512}\left(\frac{2p}{a}\right)^4(3-13\sin^2\theta-22\sin^4\theta)+\dots. \tag{48}$$

This agrees with the expansion which may be found by the same method applied directly to this case. If we make a and d infinite, the stream is bounded by a lower plane wall. In this case the coefficients have the same numerical values, but r_0 and r_2 are now negative, and we see that the result is the same as (48) but with the terms in the odd powers of $2p/a$ negative.

Another special case is when the mid-point of the plate is midway between the walls, or $a = b = \frac{1}{2}d$. In this case

$$r_0 = 0;\quad r_1 = \pi^2/4d^2;\quad r_2 = 0;\quad r_3 = \pi^4/8d^4;$$
$$r_1' = \pi^2/24d^2;\quad r_3' = \pi^4/240d^4, \tag{49}$$

and we obtain

$$L/L_0 = 1+\frac{\pi^2}{24}\left(\frac{2p}{d}\right)^2(1+\sin^2\theta)-\frac{\pi^4}{7680}\left(\frac{2p}{d}\right)^4(11-53\sin^2\theta-22\sin^4\theta)+\dots. \tag{50}$$

This agrees with the expansion given by Tomotika (1934) for this particular case.

In general, calculations may be made from (45) and (46), and the variation in lift examined as the plate is moved across the channel. The following values illustrate this for one particular case:

$\theta = 10°$; $2p/d = 0{\cdot}2$

a/d	0·3	0·4	0·5	0·6	0·7
L/L_0	1·071	1·037	1·017	1·002	0·989

8. We now obtain a similar expansion for the moment of the forces about the origin. If M_0 is the moment in an infinite stream,

$$M_0 = \pi\rho c^2p^2\sin\theta\cos\theta. \tag{51}$$

Using (35) in (14), we obtain, after some reduction,

$$M/M_0 = B_1+kpr_0B_1\sin\theta+2kp^2r_1B_2\sin^2\theta$$
$$-4p^2r_1'(kB_2-\tfrac{1}{2}B_1^2)\sin^2\theta-\tfrac{1}{8}p^3r_2\{3kB_3(\sin\theta-4\sin^3\theta)-2B_1B_2\sin\theta\}$$
$$-\tfrac{1}{8}p^4r_3(4kB_4\cos 2\theta-2B_1B_3)\sin^2\theta$$
$$+p^4r_3'(kB_4-2B_1B_3+2B_2^2)\sin^2\theta\cos 2\theta+\dots. \tag{52}$$

Substituting from (36) and (43), and collecting the various terms, we obtain

$$M/M_0 = 1 + c_1 p + c_2 p^2 + c_3 p^3 + c_4 p^4 + \dots, \tag{53}$$

$$\left.\begin{aligned}
c_1 &= 2r_0 \sin\theta, \\
c_2 &= 3r_0^2 \sin^2\theta + \tfrac{1}{2} r_1 (1 + 4\sin^2\theta) - r_1', \\
c_3 &= 4r_0^3 \sin^3\theta + r_0 r_1 (3\sin\theta + 2\sin^3\theta) \\
&\quad - \tfrac{1}{4} r_2 (\sin\theta - 8\sin^3\theta) - 2r_0 r_1' (3\sin\theta - 4\sin^3\theta), \\
c_4 &= 5r_0^4 \sin^4\theta + \tfrac{15}{2} r_0^2 r_1 \sin^2\theta - 2r_0 r_2 (\sin^2\theta - 3\sin^4\theta) \\
&\quad + \tfrac{1}{4} r_1^2 (1 + 14\sin^2\theta - 12\sin^4\theta) - \tfrac{1}{8} r_3 (1 - 8\sin^4\theta) \\
&\quad - 3r_0^2 r_1' (5\sin^2\theta - 8\sin^4\theta) - r_1 r_1' (1 + 10\sin^2\theta - 20\sin^4\theta) \\
&\quad + r_1'^2 (1 + 6\sin^2\theta - 16\sin^4\theta) + \tfrac{1}{4} r_3' (1 - 4\sin^2\theta).
\end{aligned}\right\} \tag{54}$$

When b and d are made infinite, this reduces to the expression for a semi-infinite stream with an upper plane boundary, namely

$$\begin{aligned}
M/M_0 &= 1 + \frac{1}{2}\left(\frac{2p}{a}\right)\sin\theta + \frac{1}{32}\left(\frac{2p}{a}\right)^2 (1 + 10\sin^2\theta) \\
&\quad + \frac{5}{128}\left(\frac{2p}{a}\right)^3 (\sin\theta + 4\sin^3\theta) - \frac{1}{512}\left(\frac{2p}{a}\right)^4 (1 - 14\sin^2\theta - 40\sin^4\theta) + \dots.
\end{aligned} \tag{55}$$

There is also a similar reduction for a lower plane boundary.

With $a = b = \frac{1}{2}d$, the mid-point of the plate being midway between the walls, we have, from (49),

$$\begin{aligned}
M/M_0 &= 1 + \frac{\pi^2}{48}\left(\frac{2p}{a}\right)^2 (1 + 6\sin^2\theta) \\
&\quad - \frac{\pi^4}{23040}\left(\frac{2p}{d}\right)^4 (11 - 174\sin^2\theta - 170\sin^4\theta) + \dots.
\end{aligned} \tag{56}$$

For the general case, we have (54) with the coefficients given by (46). As a numerical example, we obtain the following values:

$\theta = 10^\circ$; $2p/d = 0{\cdot}2$

a/d	0·3	0·4	0·5	0·6	0·7
M/M_0	1·059	1·030	1·010	0·994	0·977

FLAT PLATE BETWEEN FREE SURFACES

9. These results may easily be modified to give approximate expressions when the stream is bounded by parallel free surfaces. At a free surface the

resultant velocity is constant; we shall take the usual approximate form of this condition which amounts to assuming the deformation of the free surface to be small and making the tangential component of the fluid velocity constant.

Thus, instead of (2), we have the boundary conditions

$$R\frac{dw}{dz} = c; \quad z = x+ia; \quad z = x-ib. \tag{57}$$

Following out the same process as in § 2, the appropriate form is now

$$\begin{aligned}\frac{dw}{dz} = c+\Sigma\frac{i^{n+1}\,n!\,A_n}{z^{n+1}} &- \int_0^\infty \frac{F^*(\kappa)\,e^{-i\kappa z-2\kappa a} - F^*(-\kappa)\,e^{i\kappa z-2\kappa b}}{1-e^{-2\kappa d}}\,d\kappa \\ &- \int_0^\infty \frac{F(-\kappa)\,e^{-i\kappa z-2\kappa d} - F(\kappa)\,e^{i\kappa z-2\kappa d}}{1-e^{-2\kappa d}}\,d\kappa,\end{aligned} \tag{58}$$

and it may be verified directly that this form satisfies the boundary conditions (57).

It follows that the expressions for the lift and moment are now

$$Y = \rho c\Gamma - 2\pi\rho\int_0^\infty \frac{F(\kappa)\,F^*(\kappa)\,e^{-2\kappa a} - F(-\kappa)\,F^*(-\kappa)\,e^{-2\kappa b}}{1-e^{-2\kappa d}}\,d\kappa, \tag{59}$$

$$\begin{aligned}M = 2\pi\rho Ri\Bigg[cA_1 &- \int_0^\infty \frac{F'(\kappa)\,F^*(\kappa)\,e^{-2\kappa a} - F'(-\kappa)\,F^*(-\kappa)\,e^{-2\kappa b}}{1-e^{-2\kappa d}}\,d\kappa \\ &- \int_0^\infty \frac{\{F'(\kappa)\,F(-\kappa) - F'(-\kappa)\,F(\kappa)\}\,e^{-2\kappa d}}{1-e^{-2\kappa d}}\,d\kappa\Bigg].\end{aligned} \tag{60}$$

It is clear that we may write down the expansions from those in the previous sections by replacing each coefficient r_n by $-r_n$, and leaving the coefficients r'_n unaltered in sign.

Hence, instead of (45) we have,

$$L/L_0 = 1+b_1\,p+b_2\,p^2+b_3\,p^3+b_4\,p^4+\dots. \tag{61}$$

$$\left.\begin{aligned} b_1 &= -2r_0\sin\theta, \\ b_2 &= 3r_0^2\sin^2\theta - r_1 - 2r'_1\cos 2\theta, \\ b_3 &= -4r_0^3\sin^3\theta + 2r_0r_1(2\sin\theta - \sin^3\theta) \\ &\quad + r_2\sin\theta\cos 2\theta + 8r_0r'_1\sin\theta\cos 2\theta, \\ b_4 &= 5r_0^4\sin^4\theta - 3r_0^2r_1(3\sin^2\theta - 2\sin^4\theta) + \tfrac{3}{4}r_1^2 + \tfrac{3}{8}r_3\cos 2\theta \\ &\quad - \tfrac{1}{2}r_0r_2(7\sin^2\theta - 12\sin^4\theta) - 18r_0^2r'_1\sin^2\theta\cos 2\theta \\ &\quad + 3r_1r'_1\cos 2\theta + 3r_1'^2\cos^2 2\theta + \tfrac{3}{4}r'_3\cos 4\theta. \end{aligned}\right\} \tag{62}$$

With b and d infinite, this gives, for a semi-infinite stream with an upper free surface,

$$L/L_0 = 1-\frac{1}{2}\left(\frac{2p}{a}\right)\sin\theta-\frac{1}{16}\left(\frac{2p}{a}\right)^2(1-3\sin^2\theta)$$
$$+\frac{1}{32}\left(\frac{2p}{a}\right)^3(3\sin\theta-5\sin^3\theta)+\frac{1}{512}\left(\frac{2p}{a}\right)^4(6-41\sin^2\theta+46\sin^4\theta)+\dots \qquad (63)$$

With $a = b = \frac{1}{2}d$, we obtain

$$L/L_0 = 1-\frac{\pi^2}{24}\left(\frac{2p}{d}\right)^2(2-\sin^2\theta)$$
$$+\frac{\pi^4}{7680}\left(\frac{2p}{d}\right)^4(64-97\sin^2\theta+66\sin^4\theta)+\dots \qquad (64)$$

For numerical comparison, we take the same case as before, and obtain the following:

$\theta = 10°;\ 2p/d = 0{\cdot}2$

a/d	0·3	0·4	0·5	0·6	0·7
L/L_0	0·924	0·951	0·969	0·983	0·994

Similarly, for the moment, we have

$$M/M_0 = 1+c_1p+c_2p^2+c_3p^3+c_4p^4+\dots, \qquad (65)$$

$$\left.\begin{aligned}
c_1 &= -2r_0\sin\theta,\\
c_2 &= 3r_0^2\sin^2\theta-\tfrac{1}{2}r_1(1+4\sin^2\theta)-r_1',\\
c_3 &= -4r_0^3\sin^3\theta+r_0r_1(3\sin\theta+2\sin^3\theta)\\
&\quad+\tfrac{1}{4}r_2(\sin\theta-8\sin^3\theta)+2r_0r_1'(3\sin\theta-4\sin^3\theta),\\
c_4 &= 5r_0^4\sin^4\theta-\tfrac{15}{2}r_0^2r_1\sin^2\theta-2r_0r_2(\sin^2\theta-3\sin^4\theta)\\
&\quad+\tfrac{1}{4}r_1^2(1+14\sin^2\theta-12\sin^4\theta)+\tfrac{1}{8}r_3(1-8\sin^4\theta)\\
&\quad-3r_0^2r_1'(5\sin^2\theta-8\sin^4\theta)+r_1r_1'(1+10\sin^2\theta-20\sin^4\theta)\\
&\quad+r_1'^2(1+6\sin^2\theta-16\sin^4\theta)+\tfrac{1}{4}r_3'(1-4\sin^2\theta).
\end{aligned}\right\} \qquad (66)$$

With b and d infinite, we obtain

$$M/M_0 = 1-\frac{1}{2}\left(\frac{2p}{a}\right)\sin\theta-\frac{1}{32}\left(\frac{2p}{a}\right)^2(1-2\sin^2\theta)+\frac{1}{128}\left(\frac{2p}{a}\right)^3(7\sin\theta-12\sin^3\theta)$$
$$+\frac{1}{256}\left(\frac{2p}{a}\right)^4(1-8\sin^2\theta+8\sin^4\theta)+\dots, \qquad (67)$$

and with $a = b = \frac{1}{2}d$, we have

$$M/M_0 = 1 - \frac{\pi^2}{24}\left(\frac{2p}{d}\right)^2 (1 + 3\sin^2\theta)$$
$$+ \frac{\pi^4}{11520}\left(\frac{2p}{d}\right)^4 (32 + 237\sin^2\theta - 395\sin^4\theta) + \dots \qquad (68)$$

For numerical comparison with previous sections, we take the same numerical case:

$$\theta = 10^\circ;\ 2p/d = 0{\cdot}2$$

a/d	0·3	0·4	0·5	0·6	0·7
M/M_0	0·938	0·965	0·982	0·998	1·011

PLANE BOUNDARY AND FREE SURFACE

10. Although the problem is not, perhaps, of practical interest, we may note that the same method can be extended to the case when one boundary, say the lower, is a rigid plane while the other, upper, boundary is a free surface; we note, again, that for a free surface the boundary condition is taken here in an approximate form.

Considering, as in § 2, the image systems formed by successive reflexions in the two surfaces, we see that these infinite series of images now consist of terms of alternate signs; summing these series we obtain

$$\frac{dw}{dz} = c + \Sigma \frac{i^{n+1}\, n!\, A_n}{z^{n+1}} - \int_0^\infty \frac{F^*(\kappa)\, e^{-i\kappa z - 2\kappa a} + F^*(-\kappa)\, e^{i\kappa z - 2\kappa b}}{1 + e^{-2\kappa d}}\, d\kappa$$
$$+ \int_0^\infty \frac{\{F(-\kappa)\, e^{-i\kappa z} - F(\kappa)\, e^{i\kappa z}\}\, e^{-2\kappa d}}{1 + e^{-2\kappa d}}\, d\kappa. \qquad (69)$$

It may be verified directly that (69) satisfies the boundary conditions

$$\left.\begin{aligned} R\frac{dw}{dz} &= c; \quad z = x + ia, \\ I\frac{dw}{dz} &= 0; \quad z = x - ib. \end{aligned}\right\} \qquad (70)$$

The expressions for the lift and the moment are

$$Y = \rho c \Gamma - 2\pi\rho \int_0^\infty \frac{F(\kappa)\, F^*(\kappa)\, e^{-2\kappa a} + F(-\kappa)\, F^*(-\kappa)\, e^{-2\kappa b}}{1 + e^{-2\kappa d}}\, d\kappa, \qquad (71)$$

$$M = 2\pi\rho R i\Bigg[cA_1 - \int_0^\infty \frac{F'(\kappa)\, F^*(\kappa)\, e^{-2\kappa a} + F'(\kappa)\, F^*(-\kappa)\, e^{-2\kappa b}}{1 + e^{-2\kappa d}}\, d\kappa$$
$$+ \int_0^\infty \frac{\{F'(\kappa)\, F(-\kappa) - F'(-\kappa)\, F(\kappa)\}\, e^{-2\kappa d}}{1 + e^{-2\kappa d}}\, d\kappa\Bigg]. \qquad (72)$$

We use the notation

$$\left.\begin{aligned} s_n &= \int_0^\infty \frac{e^{-2va}+(-1)^n e^{-2vb}}{1+e^{-2vd}}\, v^n dv, \\ s'_n &= \int_0^\infty \frac{e^{-2vd}}{1+e^{-2vd}}\, v^n dv. \end{aligned}\right\} \qquad (73)$$

By comparison with the expressions for the flat plate between two rigid boundaries, it is easily seen that we may write down the corresponding results by replacing the coefficients r_n by $-s_n$ and r'_n by $-s'_n$.

Thus we obtain, making these changes in (45),

$$\left.\begin{aligned} L/L_0 &= 1+b_1 p+b_2 p^2+b_3 p^3+b_4 p^4+\ldots, \\ b_1 &= -2s_0 \sin\theta, \\ b_2 &= 3s_0^2 \sin^2\theta - s_1 + 2s'_1 \cos 2\theta, \\ b_3 &= -4s_0^3 \sin^3\theta + 2s_0 s_1(2\sin\theta - \sin^3\theta) + s_2 \sin\theta \cos 2\theta - 8s_0 s'_1 \sin\theta \cos 2\theta, \\ b_4 &= 5s_0^4 \sin^4\theta - 3s_0^2 s_1(3\sin^2\theta - 2\sin^4\theta) + \tfrac{3}{4}s_1^2 + \tfrac{3}{8}s_3 \cos 2\theta \\ &\quad - \tfrac{1}{2}s_0 s_2(7\sin^2\theta - 12\sin^4\theta) + 18s_0^2 s'_1 \sin^2\theta \cos 2\theta - 3s_1 s'_1 \cos 2\theta \\ &\quad + 3s_1'^2 \cos^2 2\theta - \tfrac{3}{4}s'_3 \cos 4\theta. \end{aligned}\right\} \qquad (74)$$

Similarly, for the moment,

$$\left.\begin{aligned} M/M_0 &= 1+c_1 p+c_2 p^2+c_3 p^3+c_4 p^4+\ldots, \\ c_1 &= -2s_0 \sin\theta, \\ c_2 &= 3s_0^2 \sin^2\theta - \tfrac{1}{2}s_1(1+4\sin^2\theta) + s'_1, \\ c_3 &= -4s_0^3 \sin^3\theta + s_0 s_1(3\sin\theta + 2\sin^3\theta) \\ &\quad + \tfrac{1}{4}s_2(\sin\theta - 8\sin^3\theta) - 2s_0 s'_1(3\sin\theta - 4\sin^3\theta), \\ c_4 &= 5s_0^4 \sin^4\theta - \tfrac{15}{2}s_0^2 s_1 \sin^2\theta - 2s_0 s_2(\sin^2\theta - 3\sin^4\theta) \\ &\quad + \tfrac{1}{4}s_1^2(1+14\sin^2\theta - 12\sin^4\theta) + \tfrac{1}{8}s_3(1-8\sin^4\theta) \\ &\quad + 3s_0^2 s'_1(5\sin^2\theta - 8\sin^4\theta) - s_1 s'_1(1+10\sin^2\theta - 20\sin^4\theta) \\ &\quad + s_1'^2(1+6\sin^2\theta - 16\sin^4\theta) - \tfrac{1}{4}s'_3(1-4\sin^2\theta). \end{aligned}\right\} \qquad (75)$$

From (73), the coefficients are given by

$$\left.\begin{aligned} s_0 &= \frac{\pi}{2d}\sec\alpha; \quad s_1 = \left(\frac{\pi}{2d}\right)^2 \sec\alpha\tan\alpha, \\ s_2 &= \left(\frac{\pi}{2d}\right)^3 \sec\alpha(\sec^2\alpha+\tan^2\alpha), \\ s_3 &= \left(\frac{\pi}{2d}\right)^4 \sec\alpha\tan\alpha(5\sec^2\alpha+\tan^2\alpha), \\ s_1' &= \pi^2/48d^2; \quad s_3' = 7\pi^4/1920d^4, \end{aligned}\right\} \tag{76}$$

with $\alpha = \pi(b-a)/2d$.

It may be verified that if we make b and d infinite, or a and d infinite, these expressions reduce to the former results for a semi-infinite stream bounded, respectively, by an upper free surface or by a lower rigid plane.

For the particular case, $a = b = \frac{1}{2}d$, we obtain

$$\begin{aligned} L/L_0 = 1 &- \frac{\pi}{2}\left(\frac{2p}{d}\right)\sin\theta + \frac{\pi^2}{96}\left(\frac{2p}{d}\right)^2(1+18\sin^2\theta) \\ &+ \frac{\pi^3}{192}\left(\frac{2p}{d}\right)^3(\sin\theta - 14\sin^3\theta) \\ &- \frac{\pi^4}{122880}\left(\frac{2p}{d}\right)^4(11+832\sin^2\theta-3712\sin^4\theta)+\ldots, \end{aligned} \tag{77}$$

and

$$\begin{aligned} M/M_0 = 1 &- \frac{\pi}{2}\left(\frac{2p}{d}\right)\sin\theta + \frac{\pi^2}{192}\left(\frac{2p}{d}\right)^2(1+36\sin^2\theta) \\ &- \frac{\pi^3}{256}\left(\frac{2p}{d}\right)^3(\sin\theta + 32\sin^3\theta) \\ &- \frac{\pi^4}{368640}\left(\frac{2p}{d}\right)^4(11+936\sin^2\theta-12800\sin^4\theta)+\ldots. \end{aligned} \tag{78}$$

The following numerical values may be compared with those in the previous sections:

$\theta = 10^\circ;\ 2p/d = 0{\cdot}2$

a/d	0·3	0·4	0·5	0·6	0·7
L/L_0	0·924	0·942	0·953	0·956	0·960
M/M_0	0·928	0·943	0·949	0·951	0·948

In this case the relative variation near the middle of the channel is much less than when the boundaries are of the same kind.

ELLIPTIC CYLINDER

11. The expressions for a flat plate have been obtained as limiting cases of those for an elliptic cylinder. We shall consider now the general case when the cylinder is in a stream between plane parallel walls, and we shall examine the moment of the forces; further, in order to simplify the calculations, we assume in this case that there is no circulation.

Referring to § 4, we have to determine $F(\kappa)$ from (27) and (28) with $\Gamma = 0$.

The process of approximation is carried out as before, and we record the result up to terms necessary to give the moment to the required approximation. We obtain

$$\begin{aligned} F(\kappa) = -\tfrac{1}{2}cp(\tfrac{1}{2}\kappa p B_1 e^{i\theta} + \tfrac{1}{16}\kappa^2 p^2 B_2 e^{2i\theta} - \tfrac{1}{16}\kappa^3 p^3 B_3 e^{3i\theta} \\ - \tfrac{1}{64}\kappa^4 p^4 B_4 e^{4i\theta} + \tfrac{1}{384}\kappa^5 p^5 B_5 e^{5i\theta} + \ldots), \end{aligned} \tag{79}$$

$$\left.\begin{aligned} B_1 = {} & e^{i\theta} - qe^{-i\theta} + \tfrac{1}{4}p^2 r_1(2qe^{i\theta} - q^2 e^{-i\theta} - e^{-i\theta}) \\ & - \tfrac{1}{2}p^2 r_1'(e^{3i\theta} - qe^{i\theta} - qe^{-3i\theta} + q^2 e^{-i\theta}) \\ & - \tfrac{1}{32}p^4 r_3\{2q(e^{-i\theta} + e^{3i\theta}) - (1+q^2)(e^{i\theta} + e^{-3i\theta})\} \\ & + \tfrac{1}{16}p^4 r_1^2\{(1+3q^2)e^{i\theta} - q(3+q^2)e^{-i\theta}\} \\ & + \tfrac{1}{8}p^4 r_1 r_1'\{4qe^{3i\theta} - (1+3q^2)(e^{i\theta} + e^{-3i\theta}) + 2q(1+q^2)e^{-i\theta}\} \\ & + \tfrac{1}{4}p^4 r_1'^2\{qe^{-5i\theta} - q^2 e^{-3i\theta} + q(1+q^2)e^{-i\theta} - 2q^2 e^{i\theta} + qe^{3i\theta} - e^{5i\theta}\} \\ & + \tfrac{1}{8}p^4 r_3'(qe^{-5i\theta} + q^2 e^{-3i\theta} - qe^{3i\theta} + e^{5i\theta}) + \ldots. \\ B_3 = {} & e^{i\theta} - qe^{-i\theta} + \ldots, \\ B_5 = {} & e^{i\theta} - qe^{-i\theta} + \ldots. \end{aligned}\right\} \tag{80}$$

B_2 and B_4 are of order p^3 and do not contribute to the value of the moment up to terms in p^4.

Using (79) and (80) in (14), we obtain, after some reduction,

$$\begin{aligned} M/\pi\rho c^2 p^2 \sin\theta\cos\theta = {} & 1 + p^2\{\tfrac{1}{2}qr_1 + r_1'(q - 2\cos 2\theta)\} \\ & + \tfrac{1}{16}p^4\{r_1^2(1+3q^2) + r_3(1+q^2-4\cos 2\theta) + 4r_1 r_1'(1+3q^2-8\cos 2\theta) \\ & + 4r_1'^2(3+3q^2-8q\cos 2\theta + 6\cos 4\theta) + 2r_3'(3+q^2-8q\cos 2\theta+6\cos 4\theta)\} + \ldots. \end{aligned} \tag{81}$$

In this expression, θ is the angle the major axis makes with the direction of the stream, a, b are the distances of the centre of the ellipse from the two walls, and the coefficients r are given in (46); further, if a', b' are the semi-axes of the ellipse, we have $p^2 = a'^2 - b'^2$ and $q = (a'+b')/(a'-b')$.

The moment for a flat plate in a stream between plane walls, and without circulation, has been obtained, by conformal transformation, by Tomotika (1933), who also gives an expansion for the case $a = b = \frac{1}{2}d$; it is

$$M/M_0 = 1+\frac{\pi^2}{48}\left(\frac{2p}{d}\right)^2(1+2\sin^2\theta) - \frac{\pi^4}{23040}\left(\frac{2p}{d}\right)^4(11-106\sin^2\theta-66\sin^4\theta)+\dots. \qquad (82)$$

If in the general result (81) we put $q = 1$ and use the values of the coefficients given in (46) and (49) we obtain again this particular result.

We shall use (81) to illustrate one point, namely the change in the moment when a flat plate is replaced by an elliptic cylinder whose major axis is of length equal to the width of the plate; thus we examine the effect of rounding the edges of the plate and giving it a finite thickness.

To simplify the calculation, we take the cylinder in the position given by $a = b = \frac{1}{2}d$. Then (81) gives

$$M/\pi\rho c^2 a'^2\sin\theta\cos\theta = \lambda\left[1+\frac{\pi^2\lambda}{1^2}\left(\frac{a'}{d}\right)^2(2q-\cos 2\theta) + \frac{\pi^4\lambda^2}{23040}\left(\frac{a'}{d}\right)^4\{99+168q^2-(300+44q)\cos 2\theta+33\cos 4\theta\}+\dots\right], \qquad (83)$$

where $\lambda = 1-b'^2/a'^2, \quad q = (a'+b')/(a'-b')$.

We begin with a flat plate of width $2a'$, and then keeping a' constant we increase b'; to simplify the calculations we have taken the position given by $\theta = 45°$ and the following table shows the result of the calculation for various values of the ratio a'/d.

b'/a' \ a'/d	0	0·1	0·2	0·3	0·4
0	1·0	1·0165	1·0673	1·1561	1·2885
0·05	0·9977	1·0159	1·0714	1·1690	1·3149
0·09	0·9917	1·0113	1·0717	1·1734	1·3361
0·13	0·9830	1·0038	1·0679	1·1799	1·3484
0·2	0·9600	0·9829	1·0535	1·1773	1·3640
0·5	0·7500	0·7780	0·8655	1·0227	1·2664

For an infinite stream ($a'/d = 0$), this process of increasing the ratio b'/a' with a' constant gives a moment which steadily decreases to zero when $b' = a'$. An interesting point which arises from these calculations is that in a stream of finite width, with plane walls, the moment rises to a maximum before decreasing to zero; with decreasing width of the channel, this maximum increases in amount and occurs at a higher value of the ratio b'/a'.

SUMMARY

The paper gives a new treatment of the problem of a flat plate in a stream bounded by plane parallel walls, including circulation round the plate. The plate is considered as the limiting case of the elliptic cylinder; an integral equation is obtained, whose solution by continued approximation leads to expansions for the lift and moment on the plate. The solution is modified to give similar results when the stream is bounded by parallel free surfaces, taking the condition at a free surface in an approximate form; and a further modification gives the case when one boundary of the stream is a plane wall and the other is a free surface. The problem of the elliptic cylinder in general is also considered with reference to the moment of the forces when the stream is bounded by plane walls and when there is no circulation.

REFERENCES

Tomotika, S. 1933 *Rep. Aero. Res. Inst. Tokio*, **7**, 357–93.
— 1934 *Rep. Aero. Res. Inst. Tokio*, **8**, 157–227.

Some experimental observations for longitude made by theodolite fitted with a shutter eyepiece

By J. de Graaff Hunter, C.I.E., Sc.D., F.R.S.
Leverhulme Research Fellow

(*Received* 14 *February* 1938)

1. The shutter eyepiece has been evolved to enable accurate observations to be made for longitude with a field instrument. A main source of inaccuracy in time observations, on which longitude determination depends, has always been the estimation of the time at which a star crosses a line in the eyepiece of the instrument. There are considerable personal differences, as between various observers, in this estimate, in addition to variations of an individual observer; and, further, star motion is somewhat irregular owing to erratic refraction. These difficulties have been overcome to a large extent by the use of the moving-wire micrometer, with which the observer endeavours to keep the star intersected for a considerable time, during which the instants at which the wire crosses certain points is recorded on a chronograph by electrical means. The moving-wire micrometer has been fitted to transit instruments and employed only for meridian transit observations; and it requires a good chronograph. It has not been available for use with a theodolite and is altogether too bulky in its ordinary form for such a purpose.

2. *The shutter principle.* The principle of the shutter eyepiece is quite different. In this a small central portion of the telescope field is obscured by an opaque body or shutter (*vide* fig. 1), located between the object glass and its focal plane, and as close as practicable to the latter. Its object is to expose a star for observation for an instant at a time controlled by clock or chronometer. The shutter is carried on a rocking arm which allows the shutter a lateral movement of amount slightly greater than its width. Movement of the rocking arm is caused by an electromagnet, controlled by a break-circuit chronometer or clock. Thus the exposures are in constant relationship with the time-keeper. For this purpose a system of relays has been contrived, whose functions are:

(1) to allow adjustment of the duration of shutter operation,

(2) to cause shutter operation at choice, either at each chronometer second, or at every third chronometer second.

Function (1) is obtained by means of a relay with double winding, the second winding being closed by a variable resistance up to 20 ohms; whereby the duration of shutter operation can have any value between 50 and 130 msec. Function (2) is secured by five multiple contact relays.

Suppose now that a star image is formed by the theodolite at the position ordinarily covered by the shutter, i.e. when not operated. For star observation the relay system is set to give star exposures of such an interval as 70 msec. every third second. During this interval the star does not appear to move* and its position may, for the present explanation, be considered to be that corresponding to the middle of the shutter operation time. The star position at any appearance is read in terms of a suitable scale; which is a set of horizontal lines on glass, at intervals of 0·001 in., rendered luminous by light admitted at the edge of the glass plate (see also § 7).

3. *Time comparisons.* As the shutter operation is in constant relationship with the chronometer action it is convenient to regard the middle of the shutter operation as the chronometer time; and for longitude work it is necessary to compare this with the rhythmic time signals of Rugby, Bordeaux, etc. For this the shutter arm is provided with an insulated contact (*vide* fig. 1), which is broken by the first movement of the arm and so for the duration of the shutter operation. This contact is inserted in the telephone circuit of the radio set; and causes the extinction of any signals during the time of shutter operation. To extinguish a signal completely it is obviously necessary that the duration of shutter operation should exceed that of the signal. The duration of the Rugby rhythmic signals is 100 msec.; so for the comparison of the chronometer time (regarded as middle time of shutter operation) the relay set is arranged (1) to make shutter operation each second, (2) to make duration of shutter operation somewhat greater than 100 msec., say 110 msec., by means of the appropriate setting of the variable resistance. In this way a very perfect extinction of one incoming rhythmic signal can be obtained and a very precise coincidence measure secured. The variation of shutter operation time from say 70 msec. for star work to 110 msec. for radio comparison involves a correction of $\frac{1}{2}(110-70) = 20$ msec., say; and there is the further correction of 50 msec. of opposite sign to allow for the use of the middle of the rhythmic signal. So after making allowance for sidereal time, chronometer error, error of emission of rhythmics, time of propa-

---

* The power of the eye to integrate the light impulses received over a short period is involved in this. In this connexion see *Proc. Roy. Soc.* A, **122**, **1929**, **318**, "Some experiments concerning the counting of scintillations produced by alpha particles", Charlton and Lea.

gation of signals, the G.S.T. of any star appearance in the theodolite becomes known; while the L.S.T. of this event can be calculated in terms of the observed zenith distance of the star. Thus the longitude becomes known.

4. *Star observations.* In practice the scale readings of a series of some twenty star appearances are taken. It has been found that such measurements can be made conveniently at intervals of 3 sec. After the star has been picked up it is suitably positioned in the field, and its successive appearances at progressively different positions of the scale are observed at intervals of 3 sec. and recorded. Thence a mean position, corresponding to the mean time of the appearances, is derived. This is expressed in scale units and reduced to seconds of arc by a suitable factor; and combined with collimation, refraction, vertical circle reading and bubble correction, gives the corresponding zenith distance. Then, if the latitude is known, the hour angle of the star can be computed.

It is advantageous to select stars for observation in pairs, east and west, chosen so that they can be observed at approximately the same Z.D. at a convenient time interval. Observations to such a pair are made with the vertical setting of the theodolite unchanged. If the refraction depends only on Z.D. then it is the same for both stars, as also is the collimation and vertical circle reading; and the mean Z.D.'s of the stars at their respective mean times of observation differ only by the observed scale and bubble readings.

5. *Paired observations.* To deal with such observations on a pair of stars, use is made of approximate values ϕ_a, λ_a of the latitude and longitude (positive when east), obtained as may be most convenient; the true values being denoted by $\phi_a + \partial\phi$, $\lambda_a + \partial\lambda$. The local times T/T' of the observations on the E/W stars can then be approximately expressed in terms of the rhythmic time signals, and the corresponding approximate hour angles H/H', being the difference of time and R.A., expressed in arc. Suppose that ζ_c is the common Z.D. of the two stars, neglecting bubble and scale corrections, and that Δ, A, $\delta\zeta$ are the declination, azimuth, sum of bubble and scale corrections of the east star, with corresponding dashed letters for the west star; then it is easily shown, by equating two alternative expressions for $\cos\zeta_c$, that

$$(\sin A - \sin A')\cos\phi_a \partial\lambda - (\cos A - \cos A')\,\partial\phi = Z, \tag{1}$$

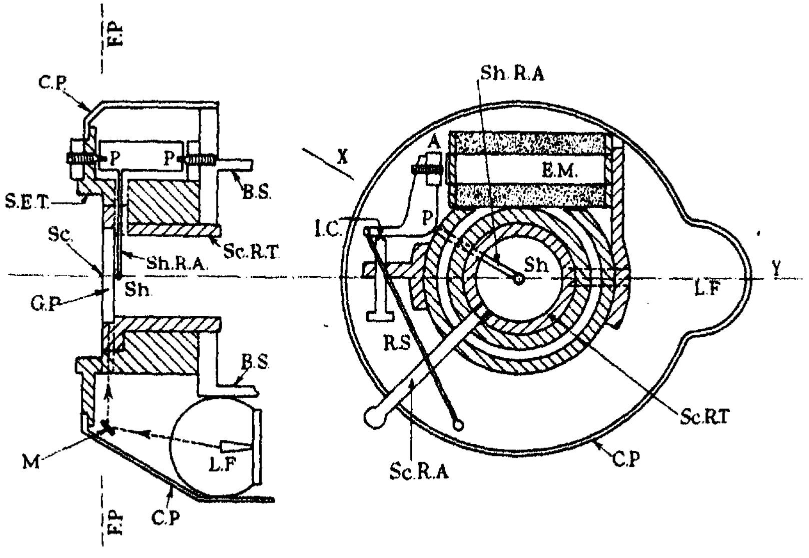
where $Z = (C - C')\operatorname{cosec}\zeta_c + \delta\zeta' - \delta\zeta$, $C = \sin\phi_a \sin\Delta + \cos\phi_a \cos\Delta \cos H$, $C' = \sin\phi_a \sin\Delta' + \cos\phi_a \cos\Delta' \cos H'$, and the longitude, λ, is reckoned positive when easterly. If x, y are the angular components of the error in assumed position, so that $x = \partial\lambda \cos\phi_a$ and $y = \partial\phi$; and if A_m is the mean azimuth of the two stars, then

$$x \cos A_m - y \sin A_m = \tfrac{1}{2}Z \operatorname{cosec} \tfrac{1}{2}(A - A'). \tag{2}$$

It is clear that from two pairs of stars, for which A_m differ considerably, two equations of form (2) serve to determine x, y. It is advantageous that $\frac{1}{2}(A - A')$ should approximate to 90°, minimizing the effect of error in Z, which depends on observation. So stars of a pair should differ in azimuth by 180° or thereabout. To find both latitude and longitude it is most favourable that the two values of A_m for two pairs, should differ by 90°. There are observational advantages in having two stars of a pair changing Z.D. at the same rate—a condition satisfied by a pair of prime vertical stars, or by a pair of meridian stars; in which case equation (2) determines x and y independently. However, considerable deviations from these ideal azimuths of stars can be admitted without serious effect on precision; and so the choice of stars is not seriously restricted. Indeed a good programme of stars can ordinarily be found without going further than the fourth magnitude.

6. *Choice of stars.* The value of the right hand side of (2) is seen to depend on a computed portion, which may be regarded as errorless, together with $(\delta\zeta - \delta\zeta') \operatorname{cosec} \frac{1}{2}(A - A')$; in which $\delta\zeta - \delta\zeta'$ depends on measurements of eyepiece scale and bubble. The bubble should be of such sensitivity as to allow of ample observational accuracy and this is practically possible. The scale readings, which indeed are differential as between the two stars of a pair, are subject to errors of observation and also of graduation. Ordinarily the readings are spread over a considerable portion of the scale; and it can be arranged that the same region is employed for both stars of a pair, in which case scale irregularities will cancel in the main. If however the star speeds in Z.D. are seriously different, this will lead to the use of more observations on one star than on the other; and the mean scale readings will have different precision. Probably the scale errors are of minor importance and then it is better to ignore matching the scale on each star while observing the same number of appearances of both. But for the optimum case of stars symmetrical to the meridian the question does not arise. The optimum observation for longitude is on a pair of stars, both in the prime vertical, for which the factor $\operatorname{cosec} \frac{1}{2}(A - A')$ is equal to unity.

The same reasoning would make the optimum observation for latitude on a pair of meridian stars, north and south of the zenith; but here there is a practical reservation. For in this case the motion in Z.D. is practically zero and successive appearances of a star would have sensibly the same scale reading. There will then be no opportunity of cancelling errors of scale—which perhaps is not important—and, apart from irregularities due to refraction, the unchanged scale reading would be subject to the well known individual error of scale estimation. So it is better to have some movement across the scale. It is believed that the irregularities due to refraction will be better smoothed out by the periodic scale readings than can be done by intersection with a micrometer wire; and that the mean of a number—20 or more—of visual estimations will lead to a more precise mean position than can be obtained by a micrometer intersection, as in the ordinary Talcott observation.



Section X—Y Plan

FIG. 1. Shutter eyepiece diagrams (on scale 1/1). *Sh.* shutter; *Sh.R.A.* shutter rocking arm; *E.M.* electromagnet; *A.* armature; *P.* armature and shutter rocking arm pivots; *I.C.* insulated contact; *R.S.* armature return spring; *L.F.* lamp filament; *Sc.R.A.* scale rotating arm; *Sc.R.T.* scale rotating tube; *C.P.* covering piece; *B.S.* breach socket, for attachment to telescope; *F.P.* focal plane; *G.P.* glass plate with scale; *Sc.* scale on eyepiece side of *G.P.*; *S.E.T.* standard eyepiece thread; *M.* mirror for illumination of *G.P.* through edge; *X*, *Y*, radial directions of the section drawing.

7. *The eyepiece scale.* In general the stars cross the scale obliquely to the vertical, so it is necessary for the lines of the scale to have considerable length. They must also be adjusted to horizontality. Some method of recognition of the lines is required and it is not practically feasible to number them on account of the small interval. The method adopted has been to omit every fifth line; except in the case of the central zero line, which

is shown by prolongations on either side. In addition there is a vertical line and the general appearance of the scale is as in fig. 2. In this are added a few letters and also a hypothetical star track *SS*. If the two star-tracks of a pair cross the centre line *AA* near the same point, which most conveniently is *C*, where the vertical line *BB* cuts *AA*, and if further the observations on the scale are equally distributed on both sides of *AA*,

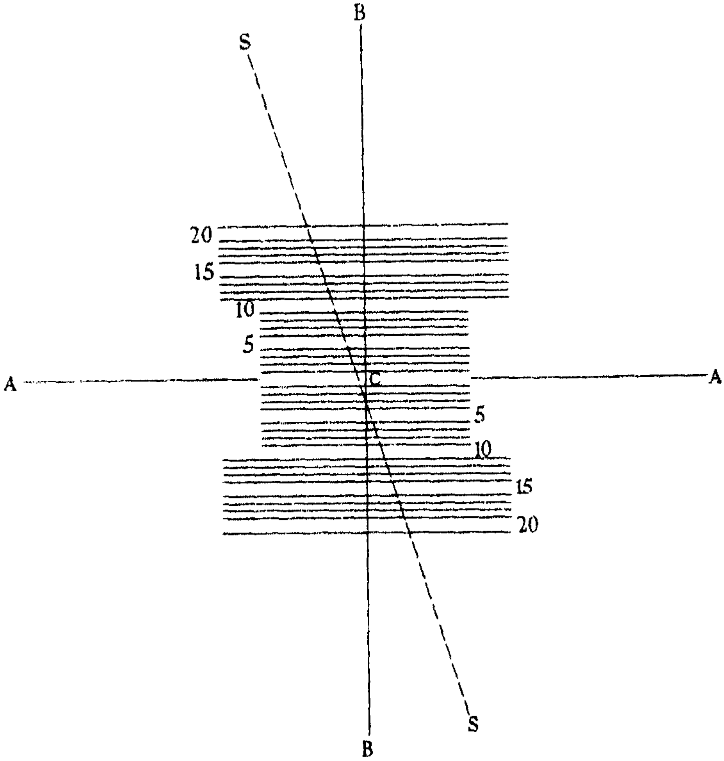


FIG. 2. Eyepiece scale (on scale 40/1).

then a small error of horizontality of the lines is unimportant. This can be arranged if the star azimuth is computed in advance and the instrument is carefully set accordingly. For observation it is most convenient to consider the scale value as increasing in the sense of the star's motion, applying later a corrective sign to one or other of a pair of stars. Readings are taken with the nearest preceding gap as zero line and so fall in groups, each of which require a gross correction to reduce to the centre zero line. An example is given in the table of § 10. If a reading of a star appearance is missed, a cross is inserted in the record so as to maintain the proper sequence. When the star track crosses the centre line AA the observer calls out "line"; and this suffices to keep all the readings in terms of the centre line. In this way the observer is spared the effort of assigning full values to each line of the scale. It is also good to record "V" at the stage when the star track crosses the vertical line BB; for then the star track is completely located, and any necessary correction for lack of horizontality can be applied. This was not actually done in the observations referred to below; in which however the horizontal adjustment was good and the star tracks all close to C. The observation is easy, for the observer soon finds his place on the scale (which remains visible all the time) and knows approximately where to expect the next appearance. It is necessary to have the appearances in terms of the chronometer time. They occur every 3 sec. As the observation time approaches the observer takes up the count of seconds and switches on the shutter action just after a second which is a multiple of 3. This results in the appearances also being at seconds which are a multiple of 3, which provides a convenient check. The seconds of the appearance preceding that which begins the sequence of scale readings is recorded and the count is then dropped.

8. *The bubble.* After the theodolite has been set for the first star of a pair and the star has appeared at the edge of the field, the bubble is looked at and a small adjustment made if needed. The bubble is read just before the chronometer time is booked. It is also read immediately after the completion of scale readings. Without touching the vertical setting, the instrument is then swung to the appropriate azimuth for the second star of the pair. In this case no further adjustment of bubble is made, as the method requires an unchanged limb setting; but otherwise the same procedure is followed for the second star. In both cases it is only the bubble reading following the scale readings which is used for computation, the earlier reading being only used as a check on the later one and on the reasonable steadiness of the theodolite. The early reading may not have allowed the bubble sufficient time to settle down to its equilibrium position. Although not required for the reduction of paired observations, it is useful to record also the constant vertical limb reading; for this can be applied in the event of the second star being missed as a pair.

The bubble reading is of great importance, entering directly into the quantity $\delta\zeta$ on which the right-hand side of (1) and (2) depend. It should be of at least as high precision as that of the mean of the eyepiece scale readings. Theodolites do not ordinarily have a vertical circle bubble of so high a precision as is desirable for this. It is to be noted that in the paired stars observation the quantity to be determined is the change of level, and so reversal of level is not called for. A bubble rigidly fixed to the instrument is accordingly suitable and better than a reversible striding level, which is open to error due to error of cross levelling as well as error due to the possibility of dust particles on the seating of the inverted V's of the striding bubble. But a great deal is demanded of the single bubble reading for one star; and so a truly sensitive bubble is needed and the bubble reading must be most carefully made. It is essential that its conditions of illumination should be good and invariable as also the relative position of the observer's eye at the moment of reading. In the observations whose results are described below, the bubble circumstances were not of the optimum character just set out; and indeed for other purposes of geodetic triangulation a bubble of less sensitivity is adequate and convenient. The precise determination of either latitude or longitude however make more severe demands on the bubble than does measurement of vertical angles for ordinary height determination. In further reference to the present observations it is also to be mentioned that no special observing pillar, insulated from the observer, was available; and the theodolite was set up on a low stand resting on hard ground. The change of bubble between observations of two stars of a pair was sometimes unduly large, and would have been unmanageable had the bubble sensitivity (1 mm. = 4″.5) been as high (1 mm. = 1″) as is desirable for the precision which could otherwise be reached.

9. *The observational equipment.* My first experiments with the shutter principle were made in 1932 with a transit instrument in India, where it has remained in use. Short references to results have been published from time to time in the Geodetic Reports of the Survey of India from 1932 onwards; and the impersonality of the method is fairly established. But the primary object has been to apply the principle to a theodolite, so as to provide the geodetic triangulator with the means of obtaining precise deviations of the vertical, without serious addition to his ordinary equipment. At the meeting of the International Union of Geodesy and Geophysics at Edinburgh in 1936 I exhibited the first shutter eyepiece fitted to a small theodolite by Messrs E. R. Watts and Son. The first opportunity of making trial observations occurred in November 1937, when I obtained some results at my house in Cottenham, Cambs. The following equipment was used:

(1) Tavistock theodolite, by Cooke, Troughton and Simms; kindly lent by the Director-General of the Ordnance Survey;

(2) Break-circuit chronometer; kindly lent by Mr Frank Mercer, St Albans;

(3) Wireless receiving set; kindly lent by the Department of Geodesy, Cambridge University;

and the following special shutter equipment:

(4) Shutter eyepiece, fitted to item (1) in place of the ordinary eyepiece;

(5) Relay box, for controlling the shutter operation;

(6) Six-volt battery, being a small car starting battery.

Of the latter three items (4) is hardly larger than the ordinary eyepiece, while (5) is a compact box weighing only a few pounds. Item (6) is accordingly the most serious increase to the impedimenta of the observer of geodetic triangulation. The precise functioning of the relays and shutter, such as will maintain the shutter lag invariable, demands a good battery of low internal resistance. Dry cells could be used but at the cost of introduction of some uncertainty in time which can hardly be estimated precisely.

A practical difficulty, which delayed the commencement of observations, arose with the method of comparing the time of shutter operation with the rhythmic time signals. The contact on the shutter arm (*vide* § 3) was introduced into the telephone circuit of the receiving set and duly interrupted the incoming signals, as had been planned. Observations however were impossible because the actual break of circuit and the relay operation introduced loud noises in the telephone. This technical difficulty was overcome by Mr B. C. Browne of the Department of Geodesy, Cambridge University. By judicious introduction of an electrolytic condenser at the relay break; insulation of both sides of the shutter contact; earthing of relay set, instrument and observer; and a transformer between radio output and telephone; Mr Browne obtained complete silence in the telephones when no signal was coming in, and only a slight scratchy sound if the break occurred during the course of a signal. For the reception of radio signals, it was arranged that the duration of shutter operation should slightly exceed the duration of the rhythmic signal. As the coincidence approaches—and indeed whenever the rhythmic signal overlaps the shutter operation—a small noise is heard. When, however, the shutter operation overlaps the rhythmic signal at both ends, then there is complete silence; and so a perfect extinction is obtained. The number of audible "scratches" enables the duration of both shutter operation and rhythmic signal to be evaluated approximately, and also gives a convenient warning of the approach of a coincidence which can easily be assessed correct to 10 msec. As the shutter movement itself is directly compared with the rhythmic signals, there is no necessity to measure the lag of the shutter in relation to the chronometer break; accordingly *no chronograph is required*. It is however necessary to reduce the result by corrections for the difference of duration of shutter operation as for radio reception and as for star work, and also from the middle to the beginning of each radio signal; with fixed rheostat settings and radio signals lasting 100 msec. these corrections combined are constant. They amount to $\frac{1}{2}$ (rhythmic duration less increase of shutter duration)—approximately 30 msec. in the present case—subtractive from east longitude. The evaluation of this quantity from the counts of extinctions and audible scratches has a maximum uncertainty of the difference of the rhythmic period from the chronometer second, i.e. 1/61 mean second or 1/73 sidereal second according to the rating of the chronometer to mean or sidereal time. The best course will be to calibrate the shutter duration in terms of the rheostat settings at the outset; by such means as the displacement of a rotating neon lamp, made luminous by the shutter current. There is also a small further correction, due to the fact that the star is actually exposed only when the shutter is open to a definite extent, differing slightly from one star to another, combined with the fact that the shutter motion is not linear in time. This correction is very small if the shutter action is suitably prompt. Some improvement in this direction in the actual shutter used is desirable, and will be made in subsequent shutters. In the present case the error can hardly reach 10 msec.

10. *The observations.* The apparatus was assembled at Cottenham on 2 November 1937. At that time weather was not good and clear skies were only occasional. I was restricted as regards time available, as it was necessary for me to leave England on 21 November. The earlier days were spent in trials as opportunity offered; and from these it was possible to plan the most convenient lay-out of the apparatus and to decide the precise sequence of the observation. My cousin, Miss Pierrepont, kindly acted as recorder and became familiar with the work in these early days, while I became familiar with the routine of the observation. Rough values of latitude and longitude were obtained. The uncertain weather made it difficult to plan a complete programme and to adhere to it; and for this reason the observations are less numerous than could be desired.

The most important matter was to obtain some estimate of the precision of the method with so small a telescope (focal length 10 in.); and this accounts for the concentration chiefly on one pair of stars, reobserved on several evenings. A second pair, observed on one evening only, served to determine the latitude; not with full precision, but sufficient to reduce the longitude work. The first pair was observed satisfactorily on 13, 14 and 15 November.

It has already been remarked that it is advantageous to observe paired stars—that is two stars of opposite aspect with unchanged vertical setting of the instrument—whereby reading and graduation errors of the vertical limb are eliminated. But it is also possible to observe these same stars before/after they reach the common zenith distance. By this process, when readings of the vertical limb are also included, it is possible to form what I call "combinations"—as opposed to "pairs"—in which the stars have nearly equal zenith distances, given by the limb readings. These can be combined to eliminate collimation error and constant errors of graduation and refraction. The results show that, in this particular set of observations, four combinations have about the same weight as one pair. The observations are distributed almost uniformly over a longer period and this may be useful in eliminating slow variations of refraction; in any case, the accordance of results of the several days observations is found to be much improved when both combinations and pairs are used.

The original plan did not contemplate the observation of combinations, and no provision had been made to observe temperature and pressure for the computation of refraction. The collimation of the centre line of the eyepiece had been measured on a target placed about 50 yd. away, for which a change of focus of the telescope was necessary. It was impossible to arrange a more distant target owing to the nature of the surroundings. Accordingly collimation and refraction had to be taken provisionally in the computations and subsequently determined from the combinations themselves. There is no objection to this course, which indeed is inherent in the method of combining two stars of nearly equal zenith distance. Those combinations which were also pairs were computed in accordance with equation (1). For the combinations each star was computed separately; but instead of computing the hour angle for each observed zenith distance, a table was constructed from which the hour angle could be interpolated.

An example of the observations is now given, to show the detail of the observations. In this group of 201 scale readings two were missed. Interpolated values, enclosed in round brackets, were used to complete the set—the most convenient practical method. None of the observed scale values were rejected with a view to "scale-matching"; so the precision of scale graduations is relied on more than need have been done. The mean value of a scale interval had been determined from a considerable number of the observations themselves to be $20''.004$, rounded off to $20''$.

11. *Reduction and results of observations.* The reduction of paired observations has been made according to equation (1), for which the reduced

TABLE. TIME OBSERVATIONS AT COTTENHAM ON 14 NOVEMBER 1937

Star ...	α Cephei (west)				ϵ Persei (east)			
Bubble ...	−0·1	0·0	−0·1	0·0	+0·1	+0·8	+1·5	−0·2
Chronometer ...	$21^{h}\,14^{m}\,0^{s}$	$21^{h}\,18^{m}\,9^{s}$	$21^{h}\,22^{m}\,33^{s}$	$21^{h}\,26^{m}\,18^{s}$	$21^{h}\,32^{m}\,12^{s}$	$21^{h}\,37^{m}\,57^{s}$	$21^{h}\,43^{m}\,3^{s}$	$21^{h}\,46^{m}\,24^{s}$
Scale readings S	2·3	1·8	0·8	1·7	2·0	3·8 −15	0·9	1·7
	3·0	2·6	2·0	2·8	3·7	0·3	2·1	3·0
	4·5 −45	3·7	3·0	4·0	4·9 −45	(1·6)	3·8 −30	4·2 −45
	0·7	4·8 −60	4·0 −60	4·4 −60	1·0	3·0	0·0	0·8
	2·0	1·3	0·0	1·0	2·7	4·2 −40	1·2	2·0
	2·9	2·0	1·0	2·0	4·0 −30	0·8	2·7	3·4
	3·7	3·0	2·0	3·0	0·3	2·0	3·9 −20	4·7 −40
	4·8 −50	4·0 −40	3·0	3·9	1·8	3·6 −15	0·3	1·2
	0·8	0·0	4·0 −50	4·9 −50	3·0	0·0	1·8	2·6
	2·1	1·2	0·0	1·0	4·3 −20	1·7	3·0	4·0 −15
	2·9	2·2	1·2	2·1	1·0	2·7	4·2 0	0·2
	4·0 −20	3·2	2·2	3·0	2·0	4·0 0	1·0	1·8
	0·0	4·1 −25	3·2	4·0	3·6 0	0·4	2·1	3·0
	1·0	0·4	4·3 −25	4·9 −25	0·2	1·8	3·2	4·0 0
	2·0	1·3	0·2	1·3	1·2	3·0	4·6 +20	0·9
	3·1	2·3	1·7	2·3	3·0	4·0 +20	1·3	(2·1)
	4·0 0	3·3	2·6	3·3	4·0 +20	0·8	2·3	3·2
	0·1	4·5 0	3·8	4·2 0	0·5	2·0	3·7 +30	4·9 +20
	1·2	0·5	4·7 0	0·4	2·0	3·5	0·3 +15	1·2
	2·3	1·8	0·8	1·4	3·2	4·9 +40		2·3
	3·5	2·9	1·7	2·4	4·3 +40	1·3		3·8 +30
	4·6 +25	3·9 +20	2·9	3·5	0·9 +15	2·3 +30		0·3 +15
	0·9	0·0	3·9 +20	4·7 +25				
	1·9	0·9	0·0	0·7				
	2·7	2·0	1·0	1·8				
	3·5	3·0	1·9	3·0				
	4·5 +50	4·0	2·9	3·8				
	0·8 +15	4·9 +60	4·0 +50	4·8 +50				
		1·0 +15	0·0	0·9 +15				
			1·1 +30					
ΣS	+44·8	+40·6	+28·9	+36·2	+33·6	+71·7	+57·4	+20·3
n	28	29	30	29	22	22	19	22
Bubble	− 0·1	0·0	0·0	0·0	0·0	+ 1·4	+ 1·8	− 0·4

Vertical	150° 31′ 20″	150° 2′ 16″	149° 31′ 20″	149° 5′ 35″	149° 5′ 35″	150° 6′ 20″	150° 43′ 1″	151° 14′ 51″
Collimation	− 4′ 0″	− 4′ 0″	− 4′ 0″	− 4′ 0″	− 4′ 0″	− 4′ 0″	− 4′ 0″	− 4′ 0″
Refraction	+ 33″.0	+ 33″.7	+ 34″.3	+ 34″.9	+ 34″.9	+ 33″.6	+ 32″.7	+ 32″.0
ζ_c				30° 50′ 59″.9	30° 50′ 59″.9			
Bubble × 4″.2	+ 0″.4	0″.0	0″.0	0″.0	0″.0	− 5″.9	− 7″.5	+ 1″.7
Scale = $(20/n)\,\Sigma S$	+ 32″.0	+ 28″.1	+ 19″.3	+ 25″.0	− 30″.5	− 1′ 5″.2	− 1′ 0″.4	− 18″.5
$\delta\zeta$				+ 25″.0	− 30″.5			
ζ_0	29° 25′ 45″.4	29° 54′ 45″.8	30° 25′ 33″.6	30° 51′ 24″.9	30° 50′ 29″.4	29° 49′ 2″.5	29° 12′ 23″.8	28° 41′ 24″.2
T_0	$21^h\,14^m\,43^s.5$	$21^h\,18^m\,54^s$	$21^h\,23^m\,19^s.5$	$21^h\,27^m\,3^s$	$21^h\,32^m\,46^s.5$	$21^h\,38^m\,31^s.5$	$21^h\,43^m\,33^s$	$21^h\,46^m\,58^s.5$

Notes for the Table. Up to the line "Vertical", the table represents the actual observations, as recorded; with following minor additions:

(1) In second and fourth columns of ε Persei scale readings, the quantities (1·6), (2·1), included in round brackets, are interpolated; actual readings being missed.

(2) The scale readings are made from the preceding scale gap, reckoned as zero. They fall in groups, ranging from zero to 4·9; and to reduce them to the common zero—indicated by the cross-line—additions of −15, −10, −5, 0, 5, 10, 15, according to the particular group, are required. Thus in the first column for α Cephei, there are three scale readings in group −15, five in group −10, four in group −5, and so on. Accordingly $-15 \times 3 = -45$; $-10 \times 5 = -50$; $-5 \times 4 = -20$, etc., are added at the end of each group to the right.

(3) The totals of scale readings, together with the group corrections, are given in line beginning ΣS. In the following line is given n, the number of scale readings.

The rest of the table shows certain reductions of the observations, to provide quantities for computation. In the Tavistock theodolite the vertical reading represents nadir distance, and so has to be subtracted from 180° for zenith distance. To zenith distance there are four corrections:

(1) Collimation. A rough value of 4 min. was used.

(2) Refraction. These correspond to standard temperature 50° and pressure 30 in. These two corrections give the approximate zenith distance.

(3) Bubble. } These two corrections combined give ζ_c, the quantity required for paired observations; and added to ζ_c give
(4) Scale. } the mean observed zenith distance, ζ_0, required for combinations.

For the bubble the factor 4·2 is used to reduce to seconds of arc.

The scale correction is $(20/n)\,\Sigma S$.

The chronometer time recorded applies to the star appearance preceding that to which the first scale reading applies. Accordingly the chronometer time T_0 corresponding to the mean scale reading is found by adding $\frac{3}{2}(n+1)$ seconds to the recorded time.

An example of paired observations is also provided in the table; where the last column of α Cephei and the first column of ε Persei were observed at the unchanged vertical setting 149° 5′ 35″.

observational data are $\delta\zeta$ and T_0. In this way the following equations were formed from observations on 13 November:

α Cephei and ε Persei $\partial\lambda'' = +0{\cdot}731\,\partial\phi - 11''{\cdot}5$

β Draconis and γ Cassiopeiae $\partial\lambda'' = -0{\cdot}232\,\partial\phi - 13''{\cdot}0$

from which it follows that $\partial\phi = -1''{\cdot}5$

and $\phi = 52^\circ\,17'\,25''{\cdot}5$.

The actual reduction of the combinations has been carried out as follows. First, each set on each star has been computed to give the hour angle and thence the longitude, corrections for $\partial\phi$ being included; and then a mean for the day for each star taken out. From the discrepancies of these means, attributable to error in zenith distance, a correction for the daily collimation and refraction has been found and applied to the results of each set. In this way the mean longitude for the group of combinations has been found for each day and also the deviations of each set therefrom. It may be mentioned that chronometer time was corrected by means of Rugby and Bordeaux rhythmic signals and the sidereal time corresponding to the approximate longitude, 32 sec. east, so derived. The corrections to the emission times of the rhythmics was received when the computations were nearly completed, and introduced at the final stage.

The mean longitude determined from all the combinations of α Cephei and ε Persei on the three evenings is 31·34 sec. east. The variations of the daily results from this mean are

13 Nov.	+0·09 sec.
14 Nov.	−0·10 sec.
15 Nov.	0·00 sec.

from which the M.S.E. for one day = ± 0·078 sec.

The variations of the results of individual sets on each star from the mean of the day are, *in units of* 0·01 *sec.*,

	α Cephei					ε Persei				
13 Nov.	+29	− 8	+25	−38	−10	−19	+20	−12	+6	+6
14 Nov.		−22	+45	+19	+50	+ 4	−20	+22	−7	
15 Nov.		0	−17	− 5	+23	−21	− 8	+24	+5	

These relatively large errors are no doubt due mainly to error of reading and graduation of the vertical limb; and they are chiefly of an accidental type. They are however considerably smaller than could be obtained without the use of the shutter eyepiece.

The mean longitude determined from the paired observations on α Cephei and ϵ Persei from three evenings is 31·39 sec. east, and the daily variations from this mean are

13 Nov. $-0{\cdot}05$ sec.
14 Nov. $+0{\cdot}11$ sec.
15 Nov. $-0{\cdot}05$ sec.

from which the M.S.E. of a pair on one day is $\pm 0{\cdot}075$ sec.

So it appears that the precision of one pair is much the same as that of four combinations. It is interesting to combine the results of pairs and combinations; and for this purpose I have ignored the fact that the observations forming pairs also enter as combinations, and also treated the combinations of any day as of the same weight as the pair of that day. The mean longitude determined from all pairs and combinations is 31·365 sec. east. The residual, v, of the combinations or pairs, as well as the residuals, V, of the daily means of combinations and pairs are, in *seconds*,

	Combinations v	Pairs v	Mean of day V
13 Nov.	$+0{\cdot}065$	$-0{\cdot}025$	$+0{\cdot}020$
14 Nov.	$-0{\cdot}125$	$+0{\cdot}135$	$+0{\cdot}005$
15 Nov.	$-0{\cdot}025$	$-0{\cdot}025$	$-0{\cdot}025$

from which

M.S.E. of one value (pair or group of combinations) = $\pm 0{\cdot}101$ sec.
M.S.E. of six values = $\pm 0{\cdot}041$ sec.
M.S.E. of one day (pair and group of combinations) = $\pm 0{\cdot}026$ sec.
M.S.E. of three days = $\pm 0{\cdot}015$ sec.

It is to be noticed that the combining of a pair with the group of combinations for one day decreases the M.S.E. considerably. The results of all the observations, which contain six values spread over three days, are reduced by the factor $15/41 = 0{\cdot}37$ when the results are grouped by days. This might be explained by supposing that there is an oscillation of the tilt of the isopycnic surfaces (surfaces of equal density of the atmosphere) embraced in the period over which the observations extend, which ranged between 28 and 35 min. on the several days. The number of observations however is too small for basing any definite conclusion of this kind.

As there is full reason to admit that the shutter method is impersonal—for this is proved by the observations in India—the values of M.S.E. found above may be considered to be free from effects of systematic error; and, in spite of the small number of observational results on which they are based, to give a fair measure of the precision of the method. They show then that, with a programme extending over less than 2 hr. and comprising say six paired stars, a M.S.E. of 0·04 sec. in longitude should be ordinarily obtainable with a theodolite of 10 in. focus having a good object glass and bubble; and better results with a longer programme. This could not be done without the shutter eyepiece, in which case errors of personality alone might amount to 0·3 sec.

The precision of longitude results naturally decreases with increase of latitude. The proper criterion is the direction of the vertical, which depends for its prime vertical component on the longitude multiplied by the cosine of the latitude. This factor is 0·61 in the present case; and this should be applied to the M.S.E.'s found above to make them comparable with equatorial observations or to express them in relation to deviations of the vertical in the prime vertical.

12. *Other applications.* Little attention has been paid to latitude determination in the above account and the observations were not planned to make a full determination of this. Manifestly the latitude can be obtained with as much precision at least as the longitude; but for that purpose star pairs close to the meridian must be observed.

Observations for azimuth should also be successful with the shutter method; and with this in view the design allows the eyepiece scale to be rotated to bring the lines vertical. The plan is to observe low altitude stars at their transit of the vertical through a convenient station of triangulation. By using low altitude stars it may be hoped that errors due to irregularities of the horizontal axis trunnions will be minimized; and in addition errors due to cross levelling will be much reduced. Further readings of the horizontal limb are in this way avoided. No observations on this plan have been made as yet.

In conclusion it may be mentioned that the shutter eyepiece can be fitted quite conveniently to a prismatic astrolabe. In this case it will be easy to distribute the observations between the two images of the star in a well balanced way. The instant at which the two images attain the same altitude would then be deducible with high precision.

13. *Acknowledgements.* I record my most grateful thanks to Mr E. S. Ritter of the G.P.O. Telephones to whom is due the entire design of the relay box which actuates the eyepiece shutter. Mr Ritter spent much time and thought on this and the result is completely satisfactory and very compact. Colonel Sir G. P. Lenox-Conyngham, F.R.S., placed the resources of the Department of Geodesy, University of Cambridge, at my disposal, and

lent the radio set. Mr B. C. Browne, M.A., Demonstrator in that Department, laid aside personal work to assist me in mine, and overcame the difficulties encountered in rhythmic signal comparisons. My cousin, Miss Ethel Pierrepont, took up the duties of recorder; without her aid it would have been impossible for me to carry out the observations. I am very much indebted to Mr Frank Mercer, St Albans, who has placed one of his admirable break-circuit chronometers at my disposal for upwards of a year. The performance of this chronometer exceeded what I had believed possible with a portable timepiece. I also thank Brigadier M. N. Macleod, D.S.O., M.C., Director-General of the Ordnance Survey, and Major M. Hotine, R.E., for helpful co-operation and interest.

The development of the subject since September 1936 has been made during my tenure of a Leverhulme Research Fellowship as part of an enquiry into the planning of Geodetic Triangulation; and I acknowledge my great indebtedness for this financial aid.

The cascade theory of electronic showers

By Dr L. Landau, *Institute for Physical Problems, Academy of Sciences, U.S.S.R.*,

and Dr G. Rumer, *Physical Institute, Academy of Sciences, U.S.S.R.*

(*Communicated by P. Kapitza, F.R.S.—Received* 18 *November* 1937)

1. Bhabha and Heitler (1937) and also Carlson and Oppenheimer (1937) have recently proposed a very ingenious theory of showers which explains nearly all the observed phenomena; both these treatments, however, contain approximations the validity of which cannot always be justified. The calculations of the first two authors make use of the method of successive approximations whose exactness is difficult to estimate, while the second authors replace one of the fundamental equations by a different one, and it is impossible to see how this replacement may distort the final result. Consequently, we have thought it useful to make the calculation in a more rigorous way; the fundamental physical ideas, however, are exactly the same as those of the authors mentioned above.

2. Consider a fast electron penetrating into a layer of matter; as long as its energy is below a certain order of magnitude, the loss of energy of the electron will be determined mainly by ionization; for higher energies, however, loss of energy by radiation (Bremsstrahlung) becomes more important. This critical energy lies in the region of about $\epsilon = 750/\bar{Z}$ million volts, where $\bar{Z}$ denotes a certain average nuclear charge (there may be several nuclei of different kinds present). Since the collision radiation, which is determined by the interaction between the electron and the nuclear field, increases as Z^2, while the ionization is approximately proportional to the number of electrons in the material, the most suitable definition of $\bar{Z}$ is

$$\bar{Z} = \Sigma_i n_i Z_i (Z_i + 1)/N, \tag{1}$$

where n_i is the number per c.c. of nuclei having the charge Z_i, and N is the total number of electrons per c.c. In the numerator we have put $Z_i(Z_i + 1)$ instead of Z_1^2, since in the case of collision radiation we have to consider not only interactions with the nuclei but also with the electrons; it is important to remember this in the case of hydrogen. This same critical energy ϵ is also that for which the absorption of photons by pair production rather than by the Compton effect begins to become important.

3. Consider a particle (electron, positron) or photon moving normally to the boundary of a sufficiently thick layer of matter. If the energy of the particle inside this layer is high enough, both photons (produced by collision radiation) and new "material particles" (produced by pair production) will appear—in other words a shower will be generated. Let us consider how the number of particles $\Pi(E)\,dE$ and the number of photons $\Gamma(E)\,dE$, with energies in the range dE and at distance x from the surface of the layer, will change in consequence of pair production and collision radiation (we assume the energy E to be so high that the Compton effect and ionization can be neglected).

Let $\gamma(E, E')dE'$ be the probability per unit path length of the production of an electron of energy between E' and $E' + dE'$ by a photon of energy E, and let $\pi(E, E')dE'$ be the probability of the radiation of a photon of energy between E' and $E' + dE'$ by an electron of energy E. The rate of change of $\Pi(E)$ caused by pair production will consequently be

$$\left(\frac{d\Pi(E)}{dx}\right)_P = 2\int_E^\infty \Gamma(u)\,\gamma(u, E)\,du. \tag{2}$$

The factor 2 occurs because there are always two particles (electron and positron) created with equal probabilities. Similarly the rate of change of $\Pi(E)$ caused by radiation will be

$$\left(\frac{d\Pi(E)}{dx}\right)_R = \int_E^\infty \Pi(u)\,\pi(u,\,u-E)\,du - \int_0^E \Pi(E)\,\pi(E,\,E-u)\,du. \tag{3}$$

The first term gives the number of particles of energy greater than E, which have lost energy by radiating a photon, thus bringing them into the energy range between E and $E+dE$; the second term gives the number of particles which leaves this interval by radiating a photon. For the rate of change of the number of photons $\Gamma(E)$ we have in a similar way

$$\left(\frac{d\Gamma(E)}{dx}\right)_P = -\int_0^E \Gamma(E)\,\gamma(E,\,u)\,du, \tag{4}$$

$$\left(\frac{d\Gamma(E)}{dx}\right)_R = \int_E^\infty \Pi(u)\,\pi(u,\,E)\,du. \tag{5}$$

Thus for the total value of the change of the numbers of material particles and photons, we shall have:

$$\frac{d\Pi(E)}{dx} = 2\int_E^\infty \Gamma(u)\,\gamma(u,\,E)\,du + \int_E^\infty \Pi(u)\,\pi(u,\,u-E)\,du - \int_0^E \Pi(E)\,\pi(E,\,E-u)\,du, \tag{6}$$

$$\frac{d\Gamma(E)}{dx} = \int_E^\infty \Pi(u)\,\pi(u,\,E)\,du - \int_0^E \Gamma(E)\,\gamma(E,\,u)\,du. \tag{7}$$

This is the basis of the cascade theory of showers.

4. The expressions for $\gamma(E, E')$ and $\pi(E, E')$ have been given by Bethe and Heitler (1934). For the region of energies which concern us here, the screening can be considered as complete and we have

$$\gamma(E, E') = A\,\frac{E'^2 + (E-E')^2 + \frac{2}{3}E'(E-E')}{E^3}, \tag{8}$$

$$\pi(E, E') = A\,\frac{E^2 + (E-E')^2 - \frac{2}{3}E(E-E')}{E^2E'}, \tag{9}$$

where
$$A = \tfrac{4}{137}(e^2/mc^2)^2\,N\bar{Z}\log 183\,\bar{Z}^{-\frac{1}{3}}\ \text{cm.}^{-1}. \tag{10}$$

Putting in numerical values we find

$$A = 1{\cdot}19 \times 10^{-26} N\bar{Z}(1 - 0{\cdot}067 \log \bar{Z}). \tag{10'}$$

It is convenient to follow Carlson and Oppenheimer (1937) and measure lengths x in terms of the dimensionless unit, so that the length is defined by

$$x = t/A. \tag{11}$$

We obtain the lengths in cm. which correspond to unit value of t for various materials, shown in Table I, together with the values of $\bar{Z}$ and ϵ.

TABLE I

	$\bar{Z}$	ϵ (volts)	Equivalent thickness corresponding to $t = 1$ cm.
Al	14	50×10^6	9·2
Pb	83	9×10^6	0·52
H_2O	7·6	100×10^6	38·0
Air	8·3	90×10^6	39·0 (height of water barometer)

The height of the atmosphere is $t = 26$. When this is done, all the material constants disappear from the fundamental equations, which then assume the form

$$\frac{d\Pi(E)}{dt} = \int_E^\infty \Pi(u)\frac{u^2 + E^2 - \frac{2}{3}uE}{u^2(u-E)}\,du - \int_0^E \Pi(E)\frac{E^2 + u^2 - \frac{2}{3}uE}{E^2(E-u)}\,du$$

$$+ \int_E^\infty \Gamma(u)\frac{2u^2 - \frac{8}{3}uE + \frac{8}{3}E^2}{u^3}\,du, \tag{6*}$$

$$\frac{d\Gamma(E)}{dt} = \int_E^\infty \Pi(u)\frac{\frac{4}{3}u^2 - \frac{4}{3}uE + E^2}{u^2E}\,du - \int_0^E \Gamma(E)\frac{\frac{4}{3}u^2 - \frac{4}{3}uE + E^2}{E^3}\,du. \tag{7*}$$

For the discussion of these equations we introduce a new independent variable s in the place of E, in the following way: if $f(E)$ is any function of E, and f_s is defined by

$$f_s = \int_0^\infty f(E)\,E^s\,dE, \quad (0 \leqslant s < \infty), \tag{12}$$

then f_s is defined uniquely in terms of $f(E)$ and the converse may also be proved. Thus to any equation valid in a complete range of E there corresponds an equation in s, and vice versa. We now transform each term of equation (6*) and (7*) to the new variables s, and also write $\xi = E/u$

$$\int_0^\infty E^s dE \int_0^\infty \Pi(u) \frac{u^2 + E^2 - \frac{2}{3}uE}{u^2(u-E)}\,du - \int_0^\infty \Pi(E)\, E^s dE \int_0^E \frac{E^2 - u^2 - \frac{2}{3}uE}{E^2(E-u)}\,du$$
$$= -\int_0^\infty \Pi(u)\, u^s du \int_0^1 \frac{(1-\xi^s)}{1-\xi}(1 - \tfrac{2}{3}\xi + \xi^2)\,d\xi,$$

$$\int_0^\infty E^s dE \int_E^\infty \Gamma(u)\, \frac{2u^2 - \frac{8}{3}uE + \frac{8}{3}E^2}{u^3}\,du = \int_0^\infty \Gamma(u)\, u^s \int_0^1 \xi^s(\tfrac{8}{3}\xi^2 - \tfrac{8}{3}\xi + 2)\,d\xi,$$

$$\int_0^\infty E^s dE \int_E^\infty \Pi(u)\, \frac{\frac{4}{3}u^2 - \frac{4}{3}uE + E^2}{u^2 E}\,du = \int_0^\infty \Pi(u)\, u^s du \int_0^1 \xi^{s-1}(\xi^2 + \tfrac{4}{3}\xi - \tfrac{4}{3})\,d\xi,$$

$$\int_0^\infty \Gamma(E)\, E^s dE \int_0^E \frac{\frac{4}{3}u^2 - \frac{4}{3}uE + E^2}{E^3}\,du = \int_0^\infty \Gamma(E)\, E^s dE \int_0^1 (\tfrac{4}{3}/\xi^2 - \tfrac{4}{3}/\xi + 1)\,d(1/\xi).$$

The definite integrals with respect to ξ occurring in these expressions, we shall denote by $A(s)$, $B(s)$, $C(s)$ and D, i.e.

$$A(s) = \int_0^1 (1-\xi^s)(1-\xi)^{-1}(1 - \tfrac{2}{3}\xi + \xi^2)\,d\xi = \tfrac{4}{3}\{\psi(s) + \gamma\} - \frac{s(5s+7)}{6(s+1)(s+2)}, \quad (13)$$

$$B(s) = \int_0^1 \xi^s(2 - \tfrac{8}{3}\xi + \tfrac{8}{3}\xi^2)\,d\xi = 2/(1+s) - \tfrac{8}{3}/(2+s) + \tfrac{8}{3}/(3+s), \quad (14)$$

$$C(s) = \int_0^1 \xi^{s-1}(\xi^2 + \tfrac{4}{3}\xi - \tfrac{4}{3})\,d\xi = \tfrac{4}{3}/s - \tfrac{4}{3}/(s+1) + 1/(s+2), \quad (15)$$

$$D = \int_0^1 (\tfrac{4}{3}\xi^2 - \tfrac{4}{3}\xi + 1)\,d(1/\xi) = 7/9, \quad (16)$$

where $\psi(s)$ is defined by $\psi(s) = \dfrac{d}{ds}\log s!$, and γ is Euler's constant.

Thus in terms of the new variable s, the system of fundamental equations assumes the simple form

$$\left.\begin{aligned} \frac{d\Pi_s}{dt} &= -A(s)\,\Pi_s + B(s)\,\Gamma_s, \\ \frac{d\Gamma_s}{dt} &= C(s)\,\Pi_s - D\Gamma_s. \end{aligned}\right\} \quad (17)$$

5. As is well known, the solutions of (17) are given by

$$\left.\begin{aligned} \Pi_s(t) &= a_s \exp(-\lambda_s t) + b_s \exp(-\mu_s t), \\ \Gamma_s(t) &= c_s \exp(-\lambda_s t) + d_s \exp(-\mu_s t), \end{aligned}\right\} \quad (18)$$

where λ_s and μ_s are the roots of the secular equations

$$\begin{vmatrix} -A(s)+x & B(s) \\ C(s) & -D+x \end{vmatrix} = 0, \tag{19}$$

so that $\lambda(s) = \lambda_s = \frac{1}{2}\{A(s)+D\} - \frac{1}{2}\sqrt{[\{A(s)-D\}^2 + 4B(s)\,C(s)]}$

$$\mu(s) = \mu_s = \tfrac{1}{2}\{A(s)+D\} + \tfrac{1}{2}\sqrt{[\{A(s)-D\}^2 + 4B(s)\,C(s)]},$$

and a_s, b_s, c_s, d_s are constants determined by the initial conditions ($t=0$). The case which interests us is when the rate of change of the number of particles is large, i.e. when the exponents are large, and consequently we consider all quantities which do not vary exponentially as being of order of magnitude unity.† With this approximation we can neglect the solutions corresponding to μ_s, and approximately satisfy the initial conditions by using only the λ_s-solutions.

An examination of the secular equation (19) shows that for change of s in the interval $0 \leqslant s \leqslant \infty$, the function $\lambda(s)$ increases from $-\infty$ to the limiting value $\frac{7}{9}$.

6. If only a single photon passes through the layer with energy E_0, the initial conditions are

$$\Pi(E) = 0, \quad \Gamma(E) = \delta(E - E_0);$$

thus to our approximation we have

$$\int_0^\infty \{\Pi(E) + \Gamma(E)\}\, E^s dE = E_0^s \exp(-\lambda_s t). \tag{20}$$

Following Bhabha and Heitler (1937) we measure the energy by means of the dimensionless quantity y, so that

$$y = \log(E_0/E) \quad \text{and} \quad \eta = \log(E_0/\epsilon), \tag{21}$$

and η will then be the quantity corresponding to the critical energy. Let $P(y,t) = \exp\phi(y,t)$ be the distribution function with respect to y so that‡

$$\{\Gamma(E) + \Pi(E)\}\, dE = -P dy. \tag{22}$$

---

† The assumption that the exponentials are respectively large or small implies that the total number of secondary particles is large. This assumption, absolutely necessary, if radiation and creation of pairs are treated separately from ionization and Compton effect, is of course made (together with other assumptions which we have avoided) by Bhabha and Heitler and by Carlson and Oppenheimer.

‡ When in the following we refer to the number of particles we mean both charged particles and light quanta.

The expression (20) then becomes

$$\int_{-\infty}^{\infty} \exp\{\phi(y,t)-sy\}\,dy = \exp(-\lambda_s t). \tag{23}$$

The integrand has a sharp maximum for a certain value of y, say $\bar{y}$; then $\bar{y}$ will be the root of the equation

$$\frac{\partial\phi(y,t)}{\partial y} = s. \tag{24}$$

Neglecting the non-exponential terms, we obtain from (23)

$$\phi(\bar{y},t)-s\bar{y} = -\lambda(s)\,t. \tag{25}$$

From equations (24) and (25), knowing the function $\phi(y,t)$, we could determine the function $\lambda(s)$; since, however, it is $\lambda(s)$ which is given, and we wish to find the function $\phi(y,t) = \log P(y,t)$, we differentiate equation (25) with respect to s, which gives

$$\frac{\partial\phi}{\partial\bar{y}}\cdot\frac{d\bar{y}}{ds}-\bar{y}-s\frac{d\bar{y}}{ds} = -\lambda'(s)\,t.$$

On account of (24) this simplifies to

$$\bar{y} = \lambda'(s)\,t. \tag{26}$$

We substitute this value $\bar{y}$ in equation (25), and write simply y instead of $\bar{y}$ in what follows.

The equations

$$\left.\begin{aligned} y &= t\lambda'(s), \\ \phi &= t\{s\lambda'(s)-\lambda(s)\}, \end{aligned}\right\} \tag{27}$$

give a parametric representation of ϕ as a function of y and t, and in fig. 1 we show the dependence of ϕ on y for the particular values $t = 10$ and $E_0/\epsilon = 100$ (for other values the scale of fig. 1 must be correspondingly altered). Since it follows from (24) that

$$\frac{\partial\phi}{\partial y} = s > 0,$$

we see then that the function increases monotonically, or in other words, the number of created particles observed in a given thickness t of the layer increases with y, i.e. with diminishing energy of these created particles.

The expression for ϕ ceases to be valid as soon as the energy of the created particles falls to the order of magnitude of the critical energy, since the ionization and Compton effect which we could neglect for higher energies, then become increasingly important. As can be seen from equations (8) and (9) photons of small energy are formed in much larger quantities, by collision radiation, than are material particles of small energy by pair production. The number of created photons with energies in the range $E', E'+dE'$ is proportional to $1/E'$. If $E' \ll \epsilon$ then the absorption of these photons is determined mainly by the Compton effect and not by pair production: consequently in this energy region the absorption coefficient is nearly proportional to $1/E'$ (as follows from the theory of the Compton effect), i.e. the numbers of photons $\Gamma(E)$ for $E<\epsilon$ is independent of the energy.

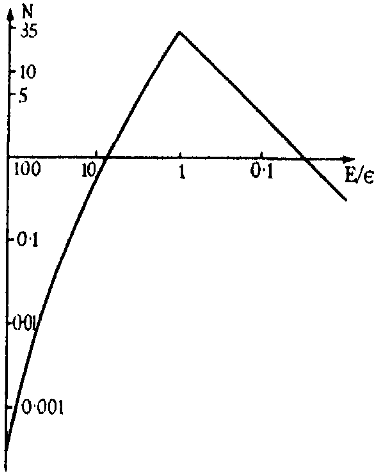


FIG. 1

As can be shown by a more detailed analysis, the same is true of the number of particles $\Pi(E)$ for $E<\epsilon'$. If we go over from E to the logarithmic measure y we find in consequence of (21) that the distribution function $P(y,t)$ for $y>\eta$ has the form

$$P(y,t) = \exp\{\text{constant} - y\}.$$

We see then that the number of particles for given penetration depth t increases with diminution of energy until the critical region is reached, when further diminution of the energy diminishes the number of particles. This means that for given penetration depth the majority of particles in a shower will have the energy ϵ.

The distribution of particles along the path of penetration into the layer of matter is given by the function which is shown graphically in fig. 2 for

$\eta = \log 100$. For other values of η the scale has to be correspondingly changed. We give in Table II the corresponding values of s, t/y, ϕ/y and λ. The curve has a maximum at the point $t = t_m$, for which the derivative $\partial\phi/\partial t$ vanishes. From (25) we have

$$d\phi - \eta\,ds - s\,d\eta = -\lambda\,dt - t\,d\lambda.$$

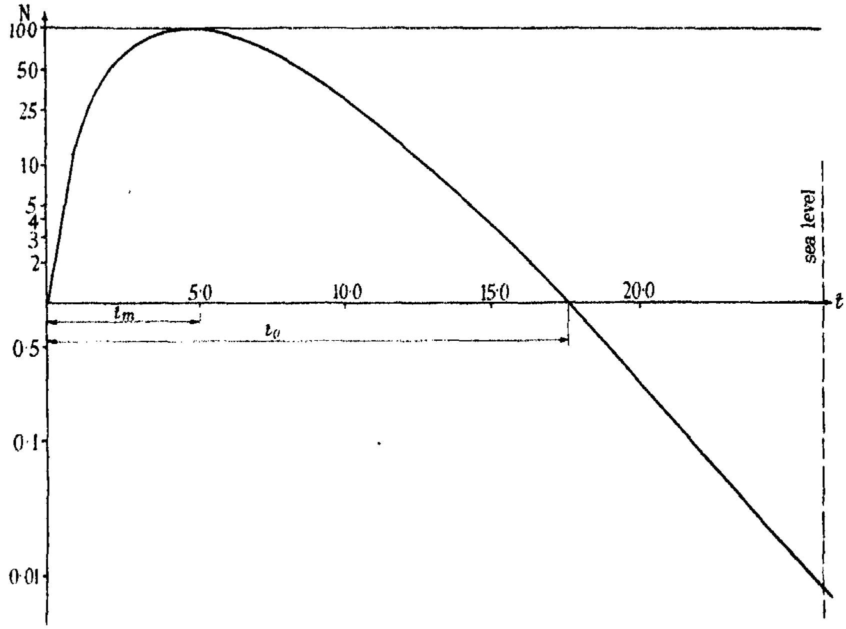


FIG. 2

Since $d\eta = 0$ and, on account of (26), $\eta\,ds = t\,d\lambda$, we obtain

$$\frac{\partial\phi}{\partial t} = -\lambda(s). \tag{28}$$

From the law of conservation of energy it follows that

$$\frac{d}{dt}\{\Gamma_1 + \Pi_1\} = 0;$$

hence $\lambda(1) = 0$; this can of course also be proved by a direct calculation.

7. Equation (26) gives for the position of the maximum

$$t_m = \eta/\lambda'(1), \quad \phi_{\max} = \eta,$$

i.e. the maximum number of particles will be observed at a depth

$$t_m = 1{\cdot}01 \log(E/\epsilon),$$

and will be equal to E_0/ϵ particles. The length of the showers track will be, expressed in t-units,

$$t_0 = 3{\cdot}8 \log(E/\epsilon),$$

which gives for air $1{\cdot}48 \log E_0/100$ m. of water barometer.

TABLE II

t/y	ϕ/y	λ	s
0·04	0·25	−3·76	0·1
0·11	0·44	−2·26	0·2
0·19	0·59	−1·56	0·3
0·28	0·71	−1·12	0·4
0·37	0·80	−1·80	0·5
0·48	0·88	−0·57	0·6
0·60	0·93	−0·38	0·7
0·72	0·97	−0·23	0·8
0·86	0·99	−0·11	0·9
1·01	1·00	0·00	1·0
1·18	0·99	0·09	1·1
1·37	0·97	0·17	1·2
1·58	0·92	0·24	1·3
1·81	0·86	0·30	1·4
2·07	0·78	0·35	1·5
2·36	0·67	0·39	1·6
2·69	0·53	0·44	1·7
3·07	0·36	0·47	1·8
3·49	0·16	0·50	1·9
3·96	−0·08	0·53	2·0
4·49	−0·37	0·55	2·1
5·08	−0·70	0·57	2·2
5·73	−1·08	0·59	2·3
6·46	−1·52	0·61	2·4
7·29	−2·03	0·62	2·5
8·17	−2·67	0·63	2·6
...	...	...	...
∞	−∞	0·78	∞

Let us calculate the fraction of the number of particles which, having penetrated the atmosphere with an initial energy of 10^{10} V, can be observed at sea-level; we have $\eta = \log 100 = 4{\cdot}6$ and $t/\eta = 26/4{\cdot}6 = 5{\cdot}65$, and from Table II we find the corresponding value of $\phi/\eta = -1{\cdot}03$ so that $\phi = -4{\cdot}74$. Consequently the required fraction is $\exp(-4{\cdot}74)$, or slightly less than 1 %. As regards the particles penetrating the atmosphere with energies of 3×10^9 V, only 1 in 1000 can reach sea-level.

In the third column of Table II we have written down the values of λ. Now equation (28) shows that λ is the coefficient of absorption and we see that the limiting value of λ (for large t) is 7/9, which gives 2 per metre of water barometer in the case of air. If we estimate the energy of the particles entering the atmosphere as of order of magnitude 3×10^9–10^{10} V, we obtain 0·6 as the order of magnitude of λ, and consequently an absorption coefficient 1·5 per metre of water barometer. This then is the size of absorption coefficient we should expect to find for light particles at sea-level.

The total energy of the shower is given by

$$E(t) = \int_0^\infty \{\Gamma(E) + \Pi(E)\}\, E\, dE = E_0 \int_{-\infty}^{\infty} \exp\{\phi(y,t) - y\}\, dy. \qquad (29)$$

The integrand has a sharp maximum at the point $y = \bar{y}$, where the derivative of the exponent vanishes, i.e.

$$\frac{\partial \phi}{\partial y} = 1 = s\left(\frac{t}{y}\right). \qquad (30)$$

As can be seen from Table II $\bar{y} \fallingdotseq t$; if $\bar{y} < \eta$, $\phi(\bar{y}, t) = \bar{y}$ and the integral becomes of order unity, and we find that the total energy of the shower is of order E_0. For $\bar{y} > \eta$ the maximum of the integrand corresponds to the point $y = \eta$ and we have

$$E(t) = E_0, \quad t < \eta$$

$$E(t) = E_0 \exp\{\phi(\eta, t) - \eta\}, \quad t > \eta.$$

The graph of this function is shown in fig. 3.

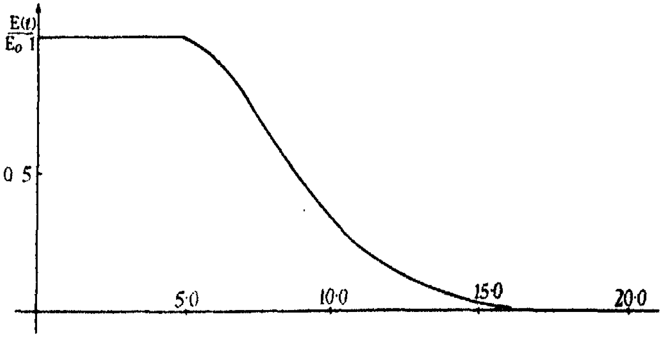


FIG. 3

8. In principle it is possible to deduce the spectral distribution of the particles entering the atmosphere and producing showers along their path

from the experimentally observable number of particles $N(t)$ per unit length of path. Let $\exp \xi(\eta)$ be the number of particles entering the atmosphere with energy E_0; each of these will produce a shower, so that the total number of particles $N(t)$ per unit length observed at depth t is

$$N(t) = \int \exp\{\xi(\eta) + \phi(\eta, t)\}\, d\eta. \tag{31}$$

The integrand has a maximum at the point $\eta = \bar{\eta}$, where the derivative $\xi'(\eta) + \dfrac{\partial \phi(\eta, t)}{\partial \eta}$ vanishes, and $\bar{\eta}$ of course depends on t. The majority of particles observed at different heights are formed by particles of different energies. In our approximation we can write

$$N(t) = \exp\{\xi(\bar{\eta}) + \phi(\bar{\eta}, t)\},$$

and we have
$$\frac{1}{N} \cdot \frac{dN}{dt} = \frac{\partial \phi}{\partial t} + \left\{\xi'(\bar{\eta}) + \frac{\partial \phi(\bar{\eta}, t)}{\partial \bar{\eta}}\right\} \frac{d\bar{\eta}}{dt} = -\lambda\left(\frac{t}{\bar{\eta}}\right). \tag{32}$$

The left-hand side of this equation can be obtained from experiment, and thus knowing $\lambda(t/\bar{\eta})$ we can determine for given depth t the corresponding value of the energy of the shower-producing particles. In the same column of Table II we find the corresponding value of $\phi(t/\bar{\eta})$, and finally obtain the required function as

$$\exp \xi(\eta) = N(t) \exp(-\phi(\bar{\eta}, t)). \tag{33}$$
$$t = t(\bar{\eta}).$$

9. We shall now consider the behaviour of cosmic particles in passing from air to a medium with a different atomic number, and first of all determine the distribution of particles with respect to energy at given height t. Let $\eta = \log(E/\epsilon)$, where E is the energy of a particle at height t, and let $\eta_0 = \log(E_0/\epsilon)$, where E_0 is the energy of a particle at the boundary of the atmosphere.

The number of particles with a given value of η will then be given by

$$N(t, \eta) = \int \exp\{\xi(\eta_0) + \phi(\eta_0 - \eta, t)\}\, d\eta_0, \tag{34}$$

or to our approximation

$$N(t, \eta) = \exp\{\xi(\bar{\eta}_0) + \phi(\bar{\eta}_0 - \eta, t)\}, \tag{34*}$$

where η_0 is the root of the equation

$$\xi'(\bar{\eta}_0) + \frac{\partial \phi(\bar{\eta}_0 - \eta, t)}{\partial \eta_0} = 0. \tag{35}$$

If we are not concerned with particles with very high energies, we can assume that η is small compared with η_0, and retain only the linear term in the expansion of the expression in the exponent in powers of η; we then have

$$\left\{\frac{\partial}{\partial\eta}\log N(t,\eta)_{\eta=0} = \{\xi'(\bar{\eta}_0)+\phi(\bar{\eta}_0-\eta,t)\}\frac{\partial\bar{\eta}_0}{\partial\eta}-\frac{\partial\phi(\bar{\eta}_0,t)}{\partial\bar{\eta}_0}\right.$$

or,
$$\log N(t,\eta) = \text{const.} - s(t/\bar{\eta}_0)\,\eta, \tag{36}$$

where the constant of course depends on t but not on η. The quantity s in this formula is connected directly with the absorption coefficient λ at the given height, and can be calculated from it by means of the table. We obtain finally

$$N(t,\eta) = N_0\exp\{-s\eta\}, \tag{37}$$

and going over to the ordinary scale of energy, we have evidently

$$dN = \text{const.}\,dE/E^{s+1} \tag{38}$$

(this formula applies of course only for $E \gg \epsilon$). For example, for $t = 10$, $E_0 = 10^{10}$ e-volts, we obtain from the table $s \sim 1{\cdot}5$, for $t = 20$, $E_0 = 10^{10}$ e-volts we find $s \sim 2{\cdot}0$.

We now determine the number of showers produced by a particle passing from air through a not too thin layer of some material; let

$$\log(\epsilon/\epsilon') = \log(\bar{Z}'/\bar{Z}) = \zeta,$$

where ϵ and $\bar{Z}$ refer to air and ϵ' and $\bar{Z}'$ to the layer, then we can write $\eta' = \eta+\zeta$. As was shown in § 7, the showers will emerge from the layer only if $\eta' > t/3{\cdot}8$ or $\eta > t/3{\cdot}8-\zeta$.

Since the number of particles diminishes with energy, for negative values of ζ (i.e. $\bar{Z}' < \bar{Z}$) we can simply write

$$\eta = t/3{\cdot}8+|\zeta| \quad \text{and} \quad N = N_0\exp\{-s(t/3{\cdot}8+|\zeta|)\}. \tag{39}$$

We see from this that with increase of the layer thickness the number of showers falls off with an absorption coefficient $s/3{\cdot}8$. For positive ζ we have to distinguish between two cases: (1) $t < 3{\cdot}8\,\zeta$ and (2) $t > 3{\cdot}8\,\zeta$; in the first case the lower limit given by the formula for η is negative, and we have simply to put $\eta = 0$, from which $N = N_0$; in the second case we obtain as before

$$N = N_0\exp\{-s(t/3{\cdot}8-\zeta)\}. \tag{40}$$

To calculate the number of particles emerging from the given layer, we have

$$N = N_0\int\exp\{-s\eta+\phi(\eta+\zeta,t)\}\,d\eta. \tag{41}$$

For negative ζ the integrand always has a maximum for a value of η different from zero, and in our approximation we can write

$$N_0 = N_0 \exp\{-s\bar{\eta} + \phi(\bar{\eta}+\zeta, t)\}, \tag{42}$$

where $\bar{\eta}$ is the root of the equation

$$s = \frac{\partial \phi(\eta+\zeta, t)}{\partial \eta} = s\left(\frac{t}{\eta+\zeta}\right). \tag{43}$$

Now we have (see equation (25))

$$\phi(\bar{\eta}+\zeta, t) - (\bar{\eta}+\zeta)s = -t\lambda(s),$$

hence
$$N = N_0 \exp\{s\zeta - t\lambda(s)\}. \tag{44}$$

In other words the number of particles continues to diminish in the material with the same absorption coefficient λ as in air. For $t = 0$,

$$N = N_0 \exp\{s\zeta\}, \tag{45}$$

i.e. since ζ is negative for a transition from air to a medium of smaller atomic number, the number of particles falls discontinuously (actually in a thin layer $t \ll 1$) by a factor $\exp s\,|\zeta|$. For a transition from air to lithium we have $\bar{Z}' = 4$, $\zeta \sim -0{\cdot}76$, which for $s \sim 1{\cdot}5$ gives a diminution of the number of particles by a factor 3.

Passing to a medium with larger Z, we have two cases according as to whether equation (43) has or has not a positive solution for $\bar{\eta}$. In the former case we evidently obtain the same result as for negative ζ, but in the latter which occurs when $s(t/\zeta) < s$ we must write

$$N = N_0 \exp \phi(\zeta, t). \tag{46}$$

This expression has a maximum for $t \cong \zeta$, as was shown, this maximum being

$$N = N_0 \exp \zeta = N_0(\bar{Z}'/Z). \tag{47}$$

We see therefore that for passage through a medium with Z larger than in air the number of particles at first increases, 10 times for lead, for instance, and then falls, reaching after a certain point the absorption coefficient for air.

10. The showers observed at great heights should evidently be ascribed to the so-called soft component of the penetrating radiation, i.e. particles produced by penetration of photons or electrons. Correspondingly the number of showers leaving a lead plate has, as we know, a maximum for a certain thickness of lead, falling off with further increase of the thickness

corresponding to the finite length of path of the showers. The number of showers at and below the earth's surface tends to a certain limit with increase of the thickness of the lead plate. As has already been pointed out, such showers can be related to secondary particles created during the passage of the hard component through matter and such particles may evidently be either electrons or photons. From the relativistic scattering formula, we can obtain for the number of fast electrons produced per unit path of the "semi-heavy" particle the following expression (we assume the velocity of the particle to be close to that of light)

$$dI = \frac{2\pi}{m} N \frac{e^4}{c^2} \frac{dE}{E^2}. \tag{48}$$

Since the number of particles increases rapidly with decrease of their energies, the number of observed showers should be very sensitive to the minimum energy of the shower which the apparatus can detect. In order to obtain a very rough estimate of the order of magnitude, we assume arbitrarily that this lower limit is $E \fallingdotseq \epsilon$ and the length of path of the shower is $t = 1$. It is seen from the calculation that the atomic number of the medium drops out of the final expression, so that the number of showers is independent of the nature of the medium, and we obtain 3 or 4 as the order of magnitude of the number of showers per 100 semi-heavy particles. Thus if the particle producing the shower is an electron we should expect that the number of showers should be independent of the mass and energy of the penetrating particle, and independent of the properties of the medium in which the shower is created, but that it should be strongly dependent on the particular experimental arrangement used for detecting the showers.

On the other hand, if it is a proton which produces the shower, calculation shows that the number of showers should depend chiefly on the mass of the penetrating particle, and also to some extent on its energy and the atomic number of the medium, but much less on the particular experimental arrangement.

The relative importance of these two possibilities depends very much on what we assume about the mass of the initial particle. Since experimentally it is found that the number of showers is independent of the medium, we may conclude that it is the first possibility which is correct.

SUMMARY

1. Starting from the physical idea proposed by Bhabha and Heitler, and Carlson and Oppenheimer, on the cascade production of showers, a more rigorous treatment of the problem is given (§§ 1–5).

2. A numerical table permits the calculation of the number of shower particles as a function of depth of penetration for any given initial energy of the creating particle, and also the energy distribution for shower particles at a given depth (§ 6).

3. The showers possess a definite range (§ 7).

4. The behaviour of a shower passing from air to a medium with a different atomic number is studied (§ 9).

5. The consequences arising from the hypothesis of the new "semi-heavy" particles are discussed.

REFERENCES

Bhabha and Heitler 1937 *Proc. Roy. Soc.* A, **159**, 432.
Bethe and Heitler 1934 *Proc. Roy. Soc.* A, **146**, 83.
Carlson and Oppenheimer 1937 *Phys. Rev.* **51**, 220.
Neddermeyer and Anderson 1937 *Phys. Rev.* **51**, 884.

Polish on metals

BY W. COCHRANE

(*Communicated by G. P. Thomson, F.R.S.—Received* 20 *January* 1938)

[Plates 12, 13]

INTRODUCTION

In the first year of this century, Beilby (1901) made microscopic and visual observations on the polished surfaces of metals. He stated that such surfaces were covered with a glass-like layer and finally reached the conclusion (Beilby 1921) that this polished layer was similar to a super-cooled liquid, the metallic atoms being in the positions which they would occupy, at a given instant, in the liquid state. His ideas did not gain general acceptance, however, amongst metallographers and other workers—thus, while Desch (1922) supports the Beilby theory, yet Tammann (1925) is definitely opposed to it. This question as to the structure of the polish layer on metals can only be decided definitely by X-rays or electron diffraction, although an experiment described by Finch, Quarrell and

Roebuck (1934) is of interest. They found that zinc atoms, evaporated on to a polished copper surface, dissolved or sank into the polish layer, showing that the latter possesses one of the properties of a liquid. There will be considerable difficulty in directing a beam of X-rays so as to be incident only on the polish layer of a specimen since this layer is so thin and since the X-rays have such a great penetrating power; one would expect that the pattern of the underlying metal would always be obtained. Cathode rays, on the other hand, will not penetrate more than a distance of about 500 A into a solid before being inelastically scattered and may therefore be used for the study of polish, when incident so as to make a small angle with the surface.

Using G. P. Thomson's electron diffraction method (Thomson 1930), French (1933) investigated several specimens of polished metals. Beginning with a rectangular block of metal, he ground one of the surfaces plane and etched it slightly. A fine pencil of cathode rays was then allowed to impinge on this surface at almost grazing incidence and the diffracted beams were recorded on a photographic plate. This gave a so-called reflexion pattern of concentric sharp rings surrounding the undiffracted central spot. Thomson (1934) has shown that this effect is completely explained by supposing that etching results in the formation, on the surface of the block, of projecting lumps through which the electrons pass. Most of the coherent scattering is due only to atoms in these projections and the dimensions of the lumps are of the order of 100 A. French then commenced to polish the surface and, on repeating the electron diffraction examination, found that the rings became less sharp until, when a good polish had been achieved, he obtained a pattern consisting of only two diffuse haloes. Subsequent researches by other workers have consistently confirmed French's result. Later, Hopkins (1935) showed that, in the case of gold, the pattern of two haloes persisted as the polish layer was removed to a depth of about 20 A and, on the gradual removal of more material, further diffuse rings of larger diameter appeared and these were all finally resolved into the sharp rings characteristic of the original non-polished polycrystalline gold. At the same time Lees (1935) obtained very similar results using copper. The halo pattern was interpreted by French as proving that the polished surface was "almost amorphous" and Hopkins deduced that this amorphous layer was approximately 20–30 A thick.

French's view was not accepted, however, by Kirchner (1932*a*, *b*), who discovered another halo effect from non-amorphous metal which has been further investigated by Papsdorf (1937). The Kirchner effect was obtained by evaporating gold (or other metals) on to thin celluloid or mica sheets so as to form metallic films 100 A thick. Using fast electrons at almost grazing incidence, these specimens yielded patterns consisting of two diffuse rings similar to those of French but, when examined by transmission, the electrons being then incident normal to the plane of the films, they gave the sharp ring patterns characteristic of polycrystalline metal of normal grain size. Kirchner showed that his results may be explained by supposing that there are on the surface of these metallic films some projecting lumps only a few atoms high. Most of the scattering at grazing incidence is caused by the atoms in these lumps and, since the number of scattering centres in each projection is very small, the pattern is diffuse, this being somewhat analogous to an optical grating consisting of only a few lines. But in transmission the sharpness of the pattern is determined by the extent of the groups of crystallographically associated atoms lying in the plane of the film, which extent is controlled solely by the size of the crystal grains themselves and the projections on the surface have no effect. Thus, on Kirchner's view, French's results might be held to indicate that the surface of the block of metal was covered, after etching, with projections of the type envisaged by Thomson. By the polishing action, these are reduced in size until they are of the dimensions of a few atoms and it is not supposed that there has been any alteration in crystal structure or crystal size.

Darbyshire and Dixit (1933) have discussed the super-cooled liquid or amorphous view and have contrasted it with Kirchner's idea that the haloes are produced merely by broadening of the sharp rings. Recently, however, Dobinski (1937) showed that previous workers had in some cases been dealing with oxidized surfaces, which are readily produced by polishing in air, thus obtaining misleading results. He was able to eliminate this defect and obtained data for clean metals. There is good agreement between his results and the theoretical values for the radii of the haloes found by postulating a liquid consisting of atoms whose closest distance of approach is the same as in the crystalline state. On the other hand, when the rings characteristic of the polycrystalline metal are considered to be broadened, taking into account their relative intensities, one also obtains haloes of the diameters yielded by the polished metal except in the case of bismuth.* This exception might be considered as deciding the

* Darbyshire and Dixit also state that Kirchner's view will not hold for body-centred cubic metals such as iron and chromium. In this they appear to be mistaken. If one takes into account the *intensities* of the various rings of the polycrystalline pattern from these metals, one finds that, on causing the rings to be broadened, the resulting pattern is in fair agreement with Darbyshire and Dixit's results for chromium and in good agreement with Dobinski's results for iron.

point at issue. The matter is more complicated, however, since Kirchner (1937) has further suggested that possibly diffuseness may be caused by deformation of the crystal lattice and also by the refractive index or inner potential of the metal coming into play. This latter view was originally put forward by Germer (1933). Not only can the inner potential cause diffuseness but it will also reduce the diameters of the rings and so add to the complexities of any argument based on measurements of these diameters. With fast electrons, however, it can only effect the results when the electrons pass through a surface at a small glancing angle (Thomson 1934).

If one were to obtain the polished metal in the form of a very thin sheet which could be examined by transmitted electrons at normal incidence, the effect of the surface projections and inner potential would then be eliminated.* Thus one would be able to decide with some certainty on the structure of the polish on a metal. The object of the present experiments was to attempt to carry out such a transmission experiment.

APPARATUS AND PROCEDURE

A piece of copper in the shape of a cylinder 0·5 cm. long and 1 cm. in diameter was obtained. One of the flat ends was ground plane on various grades of emery paper and etched for 5 min. in a 10 % solution of ammonium persulphate in water. The copper was then covered with a layer of nickel by electrolysis in a bath containing 100 g. of nickel sulphate crystals in 300 c.c. of water. The plane face of the specimen was rubbed lightly for a few strokes on grade 0000 emery paper. This gave a fairly smooth nickel surface with some light scratches. The specimen was then covered with a thin layer of gold by electrolysis. The plating bath contained 1 g. of 24 carat gold in 200 c.c. of solution. The pure gold was dissolved in *aqua regia*, the salt evaporated to dryness and then converted to the fulminate with ammonia. After washing, the fulminate was dissolved in a solution containing 4·6 g. of potassium cyanide and the whole made up to 200 c.c. with water. The anode in the plating bath was a spiral of 24 carat gold wire. The current density was found to vary considerably during the plating but was generally about 1 mA. The time of plating was 5 sec. approximately and was determined solely by judgement, the specimen being carefully watched and removed from the bath when a visible layer of gold was deposited. The specimen was then washed in water and allowed to dry in a current of air. Gold on the curved surface and bottom face of the specimen was next removed, the layer on the top surface remaining. This was examined in the electron diffraction apparatus which was of the Thomson-Fraser type (Thomson and Fraser 1930). The pattern obtained by reflexion showed sharp rings characteristic of polycrystalline gold. The specimen was removed from the apparatus and the gold surface polished for one hour on dry chamois leather. On again examining by electron diffraction, French's polish pattern of two diffuse haloes was obtained (fig. 1, Plate 12). The polished gold layer was then removed. This was done by placing the specimen, gold surface uppermost, in a photographic developing dish containing water, the specimen being just submerged. A few c.c. of a solution containing 50 % nitric acid in water were added and, in a short time, the film of gold began to be freed from the nickel. It was found that the gold broke into small pieces. The films of gold were lifted out, using fine-meshed nickel gauze, and washed by immersion in another dish containing water. The specimens were finally removed on nickel gauze and allowed to dry. They were then examined by electron diffraction using transmitted electrons.

---

* The effect of distortion is discussed later in this paper.

Nickel was used as the under layer, being a metal unlikely to alloy with gold. The condition of the nickel surface was of some importance. The surface at the end of the nickel plating process was comparatively rough and, if it was not ground slightly on 0000 emery paper, a polish could not be obtained on the gold layer. On the other hand, if the nickel was highly polished previous to gold plating, the gold deposit adhered badly and was easily rubbed off.

Experiments were also performed with chromium. A copper specimen was used of the same dimensions as above. After etching, this was plated with chromium using a bath containing 400 g. of chromic acid and 2 c.c. of sulphuric acid per litre. A fairly thick layer of chromium was obtained and, after examination in the electron diffraction apparatus, the specimen was polished on 0000 emery paper. It was verified that the polished chromium gave the correct diffuse ring pattern. The copper was dissolved away by immersing in a 50 % solution of nitric acid in water. The chromium sheet remaining was always distorted and cracked. It was washed in water and allowed to dry. The sheet was next floated on the surface of dilute hydrochloric acid, the polished side being uppermost. This caused the chromium to be thinned and the film was removed and washed, using nickel gauze as before. Only in one case was this procedure successful. The chromium appeared to be rather porous and the upper surface and edges of the film were always attacked by the hydrochloric acid so that usually the whole film was dissolved away.

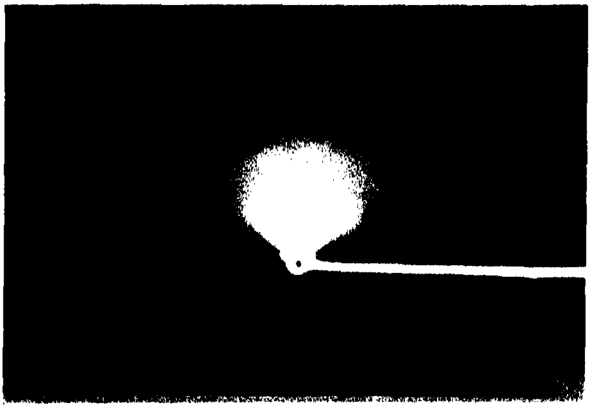


Fig. 1. Reflexion photograph from polished gold.

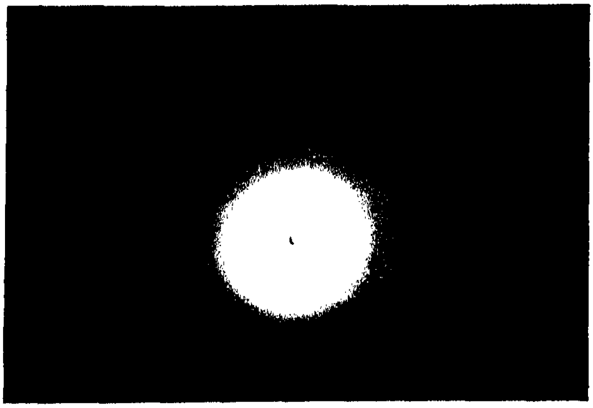


Fig. 2. Transmission photograph from polished gold.

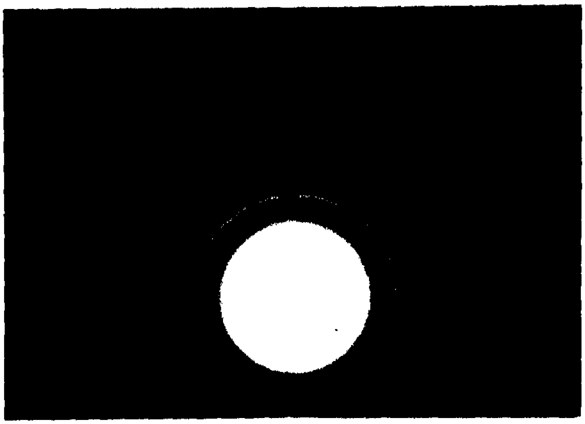


Fig. 3. Transmission photograph from polished gold after 15 hr.

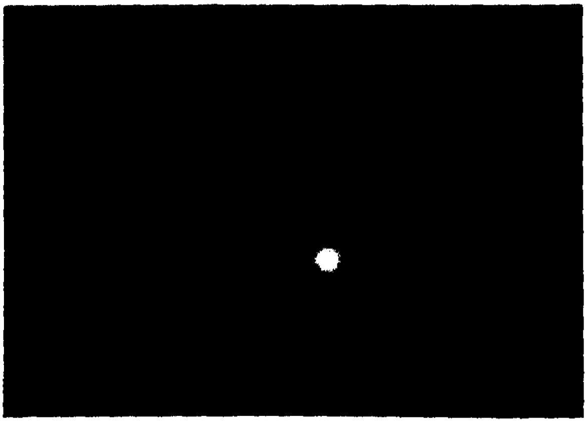


Fig. 4. Transmission photograph from spluttered gold film 300 A thick.

RESULTS

The polished gold films, when examined by transmission, yielded patterns of three diffuse rings (fig. 2, Plate 12), two being the usual polish rings and the third fainter and of larger diameter.* The Bragg spacings corresponding to these diffuse rings were 2·10, 1·30 and 0·96 A. Dobinski (1937) obtained by reflexion 2·23 and 1·27 A. There is also sign of a faint sharp ring of small diameter giving a Bragg spacing of 3·1 A. This is not a gold spacing and is attributed to the fact that the films could not be washed thoroughly owing to their being so fragile, so that some foreign substance may have been retained. On leaving the gold film in the electron diffraction apparatus overnight (15 hr.) and again examining, the pattern was completely changed and showed sharp rings characteristic of polycrystalline gold of normal grain size (fig. 3, Plate 13). The stripping of the polished gold was successfully carried out twice and this effect occurred each time.

The chromium film gave a pattern of five rings agreeing with the known structure of chromium (body-centred cubic type) and the rings appeared to be broadened compared with the "reflexion" pattern taken before polishing. The broadening was, however, slight.

TRANSMISSION THROUGH GOLD FILMS

Most transmission electron diffraction patterns obtained by various investigators have been from very thin films. It was considered necessary to test the effect of the thickness of the film on the pattern, as increasing the thickness might possibly alter the usual sharp ring pattern. Gold films of various thicknesses were prepared by spluttering on to celluloid bases, the celluloid being subsequently dissolved away in acetone and the gold mounted on nickel gauze. The thickness of these films was determined by finding the fraction of light of wave-length 4900 A transmitted and using the experimental data given by Schulze (1933). It was found that the ring patterns were still obtained with thick films until, with a film 500 A thick, no pattern was obtained. The pattern from a film 300 A is shown in fig. 4 (Plate 13).† With such thick films the rings are not so distinct, however, and the breadth of the ring due to the (200) interferences cannot be easily judged. The (220) ring remains quite noticeably sharp. The patterns from thinner films are not reproduced. They are similar to fig. 3. In these and other experiments described here, the accelerating potential difference for the electron beam was generally about 32,000 V and was measured by means of a spark-gap. The distance between the specimen and the photographic plate was 28 cm. The polished gold films were not photometered but their thickness was estimated visually to be of the order of 200–250 A.

---

* The first diffuse ring cannot be seen well on the reproduction owing to the intensity near the central spot. There is also a faint shadow of the supporting gauze in fig. 2, Plate 12 and fig. 4, Plate 13.

† Printed, using an enlarger, on ultra-hard paper so that the contrast between the rings and the background has been enhanced.

DISCUSSION

The results indicate that the polished gold layer of thickness 200–250 A yields a pattern of diffuse rings when examined by transmitted electrons incident normally on the surface. It is now proposed to attempt to interpret this result.

The experiments of the previous section show that these diffuse rings cannot be caused by the thickness of the gold films used. The effect of comparatively large thickness is principally to add to the general background of the pattern until, finally, with increasing thickness the rings disappear, being, as it were, submerged in the general diffuse scattering.

Secondly, the diffuse rings might be considered as due to distortion of the gold lattice. Up to the present, such an effect due to distortion has not been observed. Thus although many investigators have, from time to time, used beaten or rolled foils of gold and other metals for electron diffraction purposes, yet the circular rings or arcs in the resulting patterns have always been tolerably sharp. The fact that the (220) and (311) interferences are not resolved and combine to form the second diffuse ring of the polish pattern leads to the necessity of postulating a lattice distortion of at least 8 %. This would involve the presence of a large amount of latent energy in the lattice. To calculate approximately the value of this energy, it may be assumed that the ordinary elastic constants of the metal can be used, which means that the energy is taken to be a quadratic function of the atomic displacement. The curve of Fuchs (1935), giving the energy as a function of interatomic distance, shows that there is some deviation from the square law for such distortions in the case of copper. However, the order of magnitude of the result will be correct. Assuming that the compressibility of gold has the value $0{\cdot}54 \times 10^{-12}$ dyne/cm.^{2}, we find, for a uniform dilatation of 10 %, an increase of energy of $8{\cdot}3 \times 10^{10}$ erg/c.c. For shear distortion, we obtain about half this value and, for simultaneous contraction and expansion in different directions, about one-third. These amounts are very large, being indeed comparable with the sublimation energy of gold which is $4{\cdot}2 \times 10^{11}$ ergs/c.c. and thus such a distortion appears improbable. Using X-rays, Wood (1932) examined many metals which were subject to rolling. By measuring the breadth of the X-ray lines, he found that a maximum distortion was reached after which there was no further change. This maximum was generally small and always less than 1 %; in the case of some metals no broadening was detectable in the X-ray pattern. The idea that there is a saturation value for the distortion is confirmed by Taylor and Quinney (1934), who made observations of the mechanical work done on a metal specimen and the heat produced. X-ray investigations on deformed gold have been made by Boas (1937*a*, *b*). The broadening of the lines is very small. He finds a change of only 0·05 % in the Bragg spacing corresponding to the breadth of the (422) line. On the other hand, the *intensities* of the lines are altered. These X-ray results are interpreted as showing that in deformed metal all the atoms suffer a small change of a fraction of 1 % in atomic spacing. In addition, small groups of the atoms undergo larger displacements—groups situated, possibly, at the boundaries of the crystal grains or around places where a "hole" has appeared. This type of distortion also gives rise to an energy change in agreement with the saturation value found by Taylor and Quinney (1934). It appears, therefore, that when electron waves are diffracted in gold which has suffered the maximum possible distortion, the interference rings will not be noticeably broadened and the diffuse rings obtained by electron diffraction from polished gold cannot be attributed to lattice distortion.

The only hypothesis remaining is that the diffuse rings are due to the smallness of the crystal grains or to some type of "amorphous" structure. The transmission photograph (fig. 2, Plate 12) shows three diffuse rings, while, by reflexion (fig. 1, Plate 12), only two are obtained. This is, however, in agreement with Hopkins's results (Hopkins 1935) as he found that, on removing a layer of depth 30 A from a polished surface, he obtained by reflexion more than two diffuse rings and, in the transmission experiment, we are certainly dealing with a layer of much greater thickness than 30 A. If we adopt the interpretation of small grain size, the fact that the (220) and (311) interferences are merged to give the second diffuse ring shows that the dimensions of the crystal grains must be, on the average, from $2a$ to $3a$ (where a is the side of the elementary face-centred cube). It is therefore clear that the diffuse rings found by Hopkins (1935) in his various reflexion patterns were at least partly due to small grain size and his deduction from his observations, that there is an increase in grain size as the polished material is removed, appears to be reasonable. The present transmission experiments will give only an average value of the grain size over a thickness of about 200 A and the polished layer in the immediate neighbourhood of the surface must therefore consist of very small crystal grains. It is necessary also to take account of the cases of antimony and bismuth discussed by Darbyshire and Dixit (1933) and Dobinski (1937). These metals do not have a close-packed structure in the crystalline state and although, as was explained in the introduction, the interpretation of the reflexion patterns is open to some doubt, there is indication that the atoms are closely packed in the polish layer of these metals. It will thus be best to regard the polish layer on a metal as in Beilby's picture (Beilby 1921)—the atoms being in the positions they would occupy, at a given instant, in a monatomic liquid and the closest distance of approach of the atoms being assumed to be the same as in the crystalline state. Since there are interatomic forces, a completely random distribution will not be probable in a liquid, and X-ray investigations have shown the presence of some quasi-crystalline structure in liquids. Thus Prins (1931) assumes a close-packed arrangement of the atoms in a monatomic liquid and Stewart (1930) considers that the atoms congregate into small groups, each group being similar to a small crystal grain. The groups, however, are not permanent, dispersing and re-forming at different places as time proceeds. At any given time, such a liquid will not differ essentially from a conglomerate of small crystal grains, provided that there is the same type of packing in both cases.

The fact that the polished layer, when freed from the base, shows a normal polycrystalline pattern after 15 hr. recalls a similar result obtained by Kirchner (1932*c*). He found that a film of cadmium iodide, which gave a diffuse ring pattern by transmission, underwent spontaneous crystal growth when left in the electron diffraction apparatus under fore-vacuum pressure overnight. It is, however, possible that this effect may be at least partly due in our case to the action of the electron beam on the specimen. In the case of a polished gold film on nickel (not stripped off) it was found that there was no change in the diffuse ring pattern after the specimen had been in the apparatus for four days under vacuum. Bruce (1937) has studied the crystal growth in the polish layer on a gold block and finds that, after heating in a vacuum furnace for 1 hr. at 280° C., the polished surface consists of crystal grains of normal size. His observations were made by electron diffraction and also visually.

The results of the experiments on polished chromium appear to be indecisive. The edges and top surface of the film were attacked during the thinning process and the pattern obtained can only be taken as confirming slightly the more definite evidence given by gold.

The author is very much indebted to Professor G. P. Thomson for suggesting this work and for invaluable help and encouragement; also to Mr R. Bruce for advice regarding the polishing of gold.

SUMMARY

When electron diffraction patterns consisting of diffuse rings are obtained by Thomson's reflexion method, their interpretation is uncertain since such diffuseness may be caused by the geometrical configuration of the surface of the specimen. The pattern of two haloes obtained by previous workers from polished metal surfaces cannot, therefore, be held to prove the existence of an amorphous layer of metal although there is some evidence in favour of this view. In order to eliminate this difficulty, a thin layer of gold was formed on a nickel base. The gold film was polished, then stripped off and examined by transmitted electrons incident normal to the film. In this case the pattern does not depend on the form of the surface of the film. The resulting pattern consisted of three diffuse rings. It was deduced that the film consisted of very small crystal grains and that the atoms in the polish layer of a metal are arranged similarly to the atoms in a monatomic close-packed liquid at a given instant. After a period of 15 hr. the polished gold film yielded a pattern of sharp rings characteristic of polycrystalline gold, showing that crystal growth had taken place with a return to the normal crystal size.

REFERENCES

Beilby, G. T. 1901 *Rep. Brit. Ass.* p. 604.
— 1921 "Aggregation and Flow of Solids." London: Macmillan.
Boas, W. 1937*a*. *Z. Kristallogr.* A, **96**, 214–24.
— 1937*b*. *Z. Kristallogr.* A, **97**, 354–69.
Bruce, R. 1937 Unpublished observations; to appear in his Thesis (London University).
Darbyshire, J. A. and Dixit, K. R. 1933 *Phil. Mag.* (7), **16**, 961–74.
Desch, C. H. 1922 "Metallography." London: Longmans.
Dobinski, S. 1937 *Phil. Mag.* (7), **23**, 397–408.
Finch, G. I., Quarrell, A. G. and Roebuck, J. S. 1934 *Proc. Roy. Soc.* A, **145**, 676–81.
French, R. C. 1933 *Proc. Roy. Soc.* A, **140**, 637–52.
Fuchs, K. 1935 *Proc. Roy. Soc.* A, **151**, 585–602.
Germer, L. H. 1933 *Phys. Rev.* **43**, 724–6.
Hopkins, H. G. 1935 *Trans. Faraday Soc.* **31**, 1095–1101.
Kirchner, F. 1932*a* *Nature, Lond.*, **129**, 545.
— 1932*b* *Ergebn. exakt. Naturw.* **11**, 64–133.
— 1932*c* *Z. Phys.* **76**, 576–96.
— 1937 *Ann. Phys., Lpz.* (5), **28**, 21–7.

Lees, C. S. 1935 *Trans. Faraday Soc.* **31**, 1102–6.
Papsdorf, W. 1937 *Ann. Phys., Lpz.* (5), **28**, 555–68.
Prins, J. A. 1931 *Naturwissenschaften*, **19**, 435–42.
Schulze, R. 1933 *Phys. Z.* **34**, 24–38.
Stewart, G. W. 1930 *Rev. Mod. Phys.* **2**, 116–22.
Tammann, G. H. J. A. 1925 "Metallography." New York: Chemical Catalog Co.
Taylor, G. I. and Quinney, H. 1934 *Proc. Roy. Soc.* A, **143**, 307–26.
Thomson, G. P. 1930 *Proc. Roy. Soc.* A, **128**, 649–61.
— 1934 *Phil. Mag.* (7), **18**, 640–56.
Thomson, G. P. and Fraser, C. G. 1930 *Proc. Roy. Soc.* A, **128**, 641–8.
Wood, W. A. 1932 *Phil. Mag.* (7), **14**, 656–65.

The spectrum of thallium chloride

By H. G. Howell, Ph.D., *University College, Southampton*,
and N. Coulson, M.Sc., *King's College, Newcastle-upon-Tyne*

(*Communicated by W. E. Curtis, F.R.S.—Received* 22 *January* 1938)

[Plate 14]

INTRODUCTION

The results of a study of the spectrum of thallium fluoride (TlF) which have already been published (Howell 1937) contained so many points of unusual interest that it was considered desirable to extend the investigation to the other halides of thallium.

In particular it was hoped to acquire sufficient information concerning the electronic transitions in TlCl to enable a comparison to be made of the electronic levels of this molecule with those of TlF. Furthermore, certain of the vibrational constants of TlCl, as listed in such reference works as Jevons' "Report on Band Spectra" (1932) and Sponer's "Molekülspektren" (1936) seemed so abnormal as to justify a re-examination of this spectrum.

It has been studied in absorption by Butkow (1929) and by Neujmin (1934), but no work appears to have been done on the emission spectrum. Butkow found that bands appeared at about 3200 A when the temperature of the absorbing vapour was 280° C., and by raising the temperature to 530° C. he was able to develop the system to 3400 A. He also reported a continuum on the short wave edge of the system at 3106 A and another which stretched from 2475 to 2545. He measured in all 49 bands and ascribed them to the molecules $TlCl_{35}$ and $TlCl_{37}$. The analysis of the $TlCl_{35}$ system proposed by him is represented by the formula

$$\nu = 31036{\cdot}0 - 283{\cdot}7n'' + 1{\cdot}10n''^2 + 428{\cdot}5n' - 202{\cdot}6n'^2 + 50{\cdot}6n'^3 - 4{\cdot}85n'^4.$$

Four of the bands could not be fitted into his scheme, and he suggested that they might belong to a weaker and overlying system. At first sight this seems a reasonable possibility, as TlF and the halides of Ga and In all give spectra containing a doublet system of bands the members of which overlap. However, in none of these cases is there such a marked difference in intensity between the two systems as for TlCl. The large anharmonic factor would indicate a rapid convergence of the vibrational levels of the upper state of the band system and he assumed that the 3106 continuum was due to this dissociation process. In the present paper which describes both the emission and the absorption spectra it will be shown that Butkow's analysis is incorrect and that there is no need to assume such a large value of $x_e \omega_e$.

Neujmin reported that diffuse bands accompanied both the band system and the 2510 continuum, the number of bands increasing with temperature. He measured 13 on the red edge of the continuum and 4 on the violet edge of the band system and interpreted them as being due to transitions from successive v'' levels to two unstable electronic levels.

EXPERIMENTAL

The experimental method for developing both the emission and absorption spectra of such a molecule as TlCl has already been described (Howell 1937). The spectrum was surveyed from 1900 to 9000 A and the bands were photographed on a Hilger E 1 spectrograph having a dispersion of 6 A/mm. at 3200 A as compared with that of 17 A/mm. used by Butkow. The continua and diffuse bands were studied with a Hilger small quartz spectrograph. Iron arc lines served as wave length standards for the E 1 plates whilst Al arc lines were used to measure the diffuse bands. Excluding isotopic heads and diffuse bands, 58 bands were measured.

DESCRIPTION OF THE SPECTRUM

The general appearance of the absorption spectrum may be seen from the accompanying plate. At a temperature of 180° C. a continuum appears having a maximum at 2520 and a red edge at 2530. At 200° C. bands appear at 3220 together with a narrow continuum at 3110, both these continua being recorded by Butkow. Absorption also begins to take place in the 1900 region and extends to higher wave lengths as the temperature rises. Between

350 and 400° C. diffuse bands develop on the low-frequency side of both continua, these being the bands studied by Neujmin. At about 450° C. another narrow continuum, hitherto unrecorded, appears at 2890 and at 500° C. the band system has extended to 3470 and the extreme ultra-violet continuum has merged into that at 2530 which has itself joined on to that at 2890.

The plate also shows higher dispersion spectrograms of both emission and absorption spectra between 3150 and 3500. As was the case with TlF the absorption is to a great extent a reversal of the emission. The bands are all degraded to the red and nearly all show a doubling at the head, the constant intensity ratio of the doublets suggesting an isotope effect. The emission spectrum consists only of the main band system, there being no sign of continua or diffuse bands even on over-exposed plates.

Another feature common to this spectrum and that of TlF is the occurrence of "transmission" lines in the absorption bands and "absorption" lines in the emission spectrum.

VIBRATIONAL ANALYSIS

Intensity considerations point to the group of heads at 3200 A as being the $\Delta v = 0$ sequence. Each member of this sequence has an accompanying head which is considered to be a Q head. On this basis the quantum scheme given in Table I has been built up. Although it is only in this first sequence that Q heads are visible, yet the bands in the other sequences show gaps in the band structure similar to those observed in TlF (the "transmission" and "absorption" lines already referred to), and the same interpretation is given to them here, viz. that they mark the position of the band origins. Actually it is doubtful if the bands of the first sequence have true Q heads. They certainly have R heads and Q or (and) P branches and a distinct region of missing lines at the origin, and it is this dividing region that gives the structure a head-like appearance.

Measurements made on both origins and R heads are included in Table I, and as was found to be the case with TlF, ΔG values based upon origins are more consistent than those calculated from R head data. All the $TlCl_{35}$ bands measured have been fitted into this scheme. The following quantum formula represents the wave numbers of the band origins:

$$\nu = 31054{\cdot}2 - 287{\cdot}47(v''+\tfrac{1}{2}) + 1{\cdot}24(v''+\tfrac{1}{2})^2 + 216{\cdot}91(v'+\tfrac{1}{2}) - 6{\cdot}80(v'+\tfrac{1}{2})^2.$$

The O–C values which are given in Table IV range from $-2{\cdot}3$ to $1{\cdot}1$ and their algebraic sum is $-0{\cdot}6$.

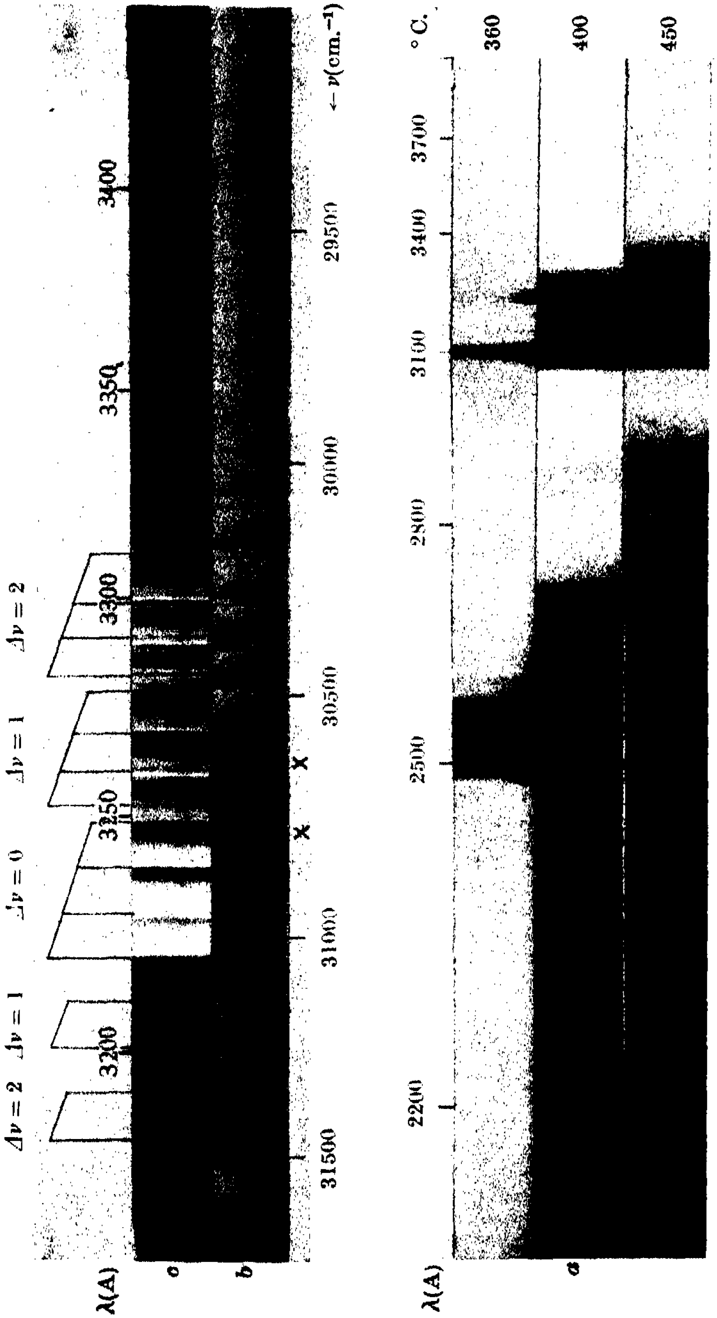


FIG. 3. The TlCl spectrum. (*a*) Absorption at various temperatures; (*b*) absorption band spectrum; (*c*) emission band spectrum.

The frequency 287·47, which is that of the ground state, since the system can be obtained in absorption, is not sensibly different from that given by Butkow, viz. 284·8, if it be remembered that his value refers to *R* heads, but ω'_e is very different, being 216·91 as against 671·5. The value proposed here is more reasonable as it is smaller than ω''_e as is the case with TlF. The actual values of the ratio ω''_e/ω'_e for TlF are 1·18, 1·32 and 1·3, whilst that proposed for TlCl is 1·32. Further, there is no need to postulate abnormal anharmonicity in the upper state as $x'_e\,\omega'_e$ is quite of the usual order of magnitude. Another feature which supports the present analysis is the variation of the separation of *R* head and origin with *v* shown in Table II. This increases as v'' increases and decreases as v' increases, which is to be expected for red-degraded bands.

The reason for the marked difference between this and Butkow's analysis is that Butkow assumed the *R* and *Q* heads of the 0, 0 band to be the *R* heads of the 0, 0 and 1, 1 bands, i.e. he considered the *R*–*Q* branch separation to be a sequence difference.

ISOTOPE EFFECT

The abundance ratios of the Cl and Tl isotopes are $Cl_{35}/Cl_{37} = 100/31{\cdot}4$ and $Tl_{205}/Tl_{203} = 100/41{\cdot}6$, and the molecules $Tl_{205}Cl_{35}$, $Tl_{205}Cl_{37}$, $Tl_{203}Cl_{35}$ and $Tl_{203}Cl_{37}$ will therefore be of relative abundance 100 : 31·4 : 41·6 : 13·0. The factor ρ representing the ratio of the reduced masses in terms of the most abundant molecule $Tl_{205}Cl_{35}$ has the following values: 0·9769, 1·0007 and 0·9776 respectively.

For a given band the isotope separation is proportional to $\rho - 1$, and the general isotopic pattern is shown in fig. 1 for a band on the long-wave side of the system origin.

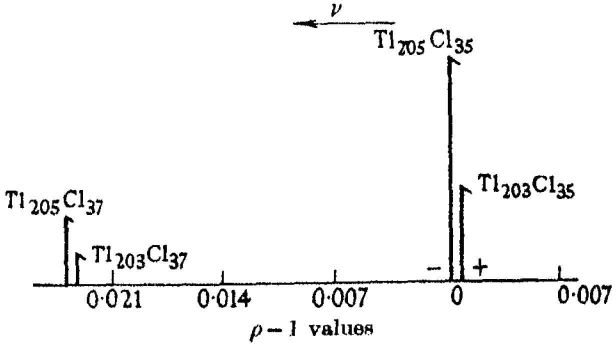


FIG. 1. Isotopic pattern for TlCl.

TABLE I. VIBRATIONAL ANALYSIS OF TlCl BAND SYSTEM

v' \ v''		0		1		2		3		4		5		6		7
0	*R*	31038·5	285·5	30753·0	279·2	30473·8	280·2	30193·6*								
	O	31016·6	286·4	30730·2	279·8	30450·4	279·8	30170·6								
	RR	189·3		194·1		193·1		196·4								
	OO	205·4		205·3		202·2		203·4								
1	*R*	31227·8	280·7	30947·1	280·2	30666·9	276·9	30390·0	274·5	30115·5	273·2	29842·3				29393·4
	O	31222·0	286·5	30935·5	282·9	30652·6	278·6	30374·0	277·7	30096·3	274·8	29821·5				
	RR	183·6		182·6		182·5		182·0		181·3		179·8				182·6
	OO	—		188·9		190·0		189·6		189·8		189·3				
2	*R*	31411·4	281·7	31129·7	280·3	30849·4	277·4	30572·0	280·2	30296·8	274·7	30022·1	272·2	29749·9	273·9	29476·0
	O			31124·4	281·8	30842·6	279·0	30563·6	277·5	30286·1	275·3	30010·8	273·1	29737·7		
				176·2				176·1		169·3		171·5		169·3		171·2
3	*R*			31305·9				30748·1	282·0	30466·1	272·5	30193·6*	274·4	29919·2	272·0	29647·2
				165·6				165·0		166·1		159·4		160·8		160·0
4	*R*			31471·5	280·6	31190·9	277·8	30913·1	280·9	30632·2	279·2	30353·0	273·0	30080·0	272·8	29807·2
										144·9						148·7
5	*R*									30777·1		—		—		29955·9

v' \ v''		7		8		9		10		11		12		13		14
0	*R*															
	O															
	RR															
	OO															
1	*R*	29393·4	275·8	29017·6	268·7	28748·9										
	O															
	RR	182·6		184·5		183·2										
	OO															
2	*R*	29476·0	273·9	29202·1	270·0	28932·1	263·9	28668·2								
	O															
		171·2		173·1		173·3		171·4								
3	*R*	29647·2	272·0	29375·2	269·8	29105·4	265·8	28839·6	262·1	28577·5						
		160·0		162·3		164·2		164·6		163·0						
4	*R*	29807·2	269·7	29537·5	267·9	29269·6	265·4	29004·2	263·7	28740·5	265·5	28475	262	28213	254	27959
		148·7		147·1		146·8		146·5		143·2		151		146		140
5	*R*	29955·9	271·3	29684·6	268·2	29416·4†	265·7	29150·7‡	267·0	28883·7	257·7	28626·0	267	28359	260	28099

* Allocation doubtful. † Overlaps 3, 8 R^i. ‡ Overlaps 3, 9 R^i.

TABLE II. SEPARATION OF *R* HEAD AND ORIGIN

v' \ v''	0	1	2	3	4	5	6
0	21·9	22·8	23·4				
1	5·8	11·6	14·3	16·0	19·2	20·8	
2		5·3	6·8	8·4	10·7	11·3	12·2

The separation of the Tl isotopes is only 1/33 that of the Cl isotopes and is too small in the present work for them to be resolved. The smallest Cl isotope separation measured is 2·6 cm.$^{-1}$ and a Tl isotope separation of the same size would correspond to a Cl separation of 86 cm.$^{-1}$, whereas the largest actually measured is only 69·6 cm.$^{-1}$.

The Cl isotope effect has been calculated from the exact relation

$$\nu^i - \nu = (\rho - 1)(\omega_e' u' - \omega_e'' u'') - (\rho^2 - 1)(x_e' \omega_e' u'^2 - x_e'' \omega_e'' u''^2),$$

where $u = v + \frac{1}{2}$.

As the measurements have been made upon R heads the following R head formula has been derived for the $Tl_{205}Cl_{35}$ heads and the appropriate constants used in the above expression:

$$\nu_{(R)} = 31079{\cdot}6 + 204{\cdot}0u' - 5{\cdot}15u'^2 - 284{\cdot}8u'' + 0{\cdot}9u''^2.$$

The agreement of the separations so calculated with those measured for 38 bands is such that the average O–C value is less than 1 cm.$^{-1}$ and is strong support for the proposed analysis, as a change of ± 1 in either v' or v'' would change the theoretical separation by a much larger quantity.

Usually the separation $\nu^i - \nu$ increases almost linearly with the distance from the system origin. However, those bands marked with a cross on the plate and which only appear in absorption provide an exception to this "rule", for their isotope effects are visibly larger than those of the 3, 3 or 1, 2 bands which are in between them. These bands fit into the 4, 4 and 5, 4 positions in the quantum scheme, and their calculated and observed isotope effects are in such excellent agreement as to justify this allocation of quantum numbers. That the effect for the 5, 4 band appears to be too large is due to the fact that the squared term in the isotope separation expression, normally small in value, here contributes more than the first power factor, thus making the approximate linear relation inapplicable.

THE ELECTRONIC STATES OF TlCl

The close resemblance of the spectrum to that of TlF suggests that the electronic transitions are similar in the two cases, and it is therefore very probable that certain, if not all, of the electronic states of TlCl may be identified by analogy. For this reason the energy diagram of TlF is reproduced in fig. 2. A comparison of the two spectra shows that:

(1) Both have an extreme ultra-violet continuum.

(2) Whereas TlF has a band system and a continuum around 2200 A, TlCl has a continuum and diffuse bands around 2500 A.

(3) Both have a weak continuum near the main bands.

(4) TlCl has only a singlet band system, whilst the corresponding TlF system is doublet.

(5) A continuum and diffuse bands accompany the band system in TlCl.

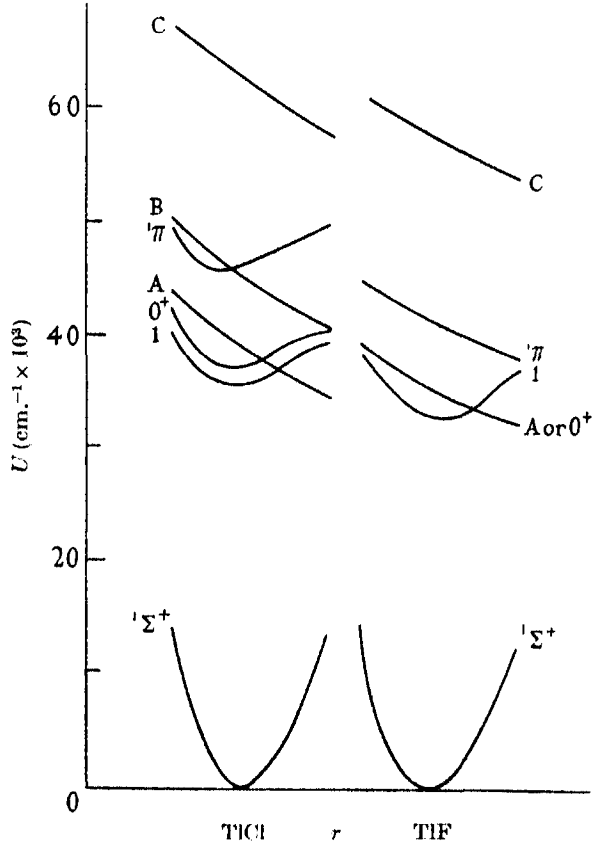


FIG. 2. Energy levels of TlF and TlCl. (The r values are quite arbitrary.)

Examination of Table I shows that no bands associated with v' levels higher than 5 are observed, suggesting that predissociation takes place above this level. The existence of an unstable level cutting the upper state of the band system is borne out by the fact that diffuse bands and continuum appear on the short-wave side of the bands. The long-wave edge of the continuum is at 3122 A, and consequently a transition from $v'' = 0$ in absorption to this unstable level involves an approximate energy change of 32000 cm.$^{-1}$. A transition from $v'' = 0$ to $v' = 5$ (the highest vibrational level) corresponds to an energy change of 31900 cm.$^{-1}$, thus making it almost certain that it is this unstable level which cuts the upper level at $v' = 5$.

The doublet TlF system was considered to be due to $0^+1 \rightleftarrows {}^1\Sigma^+$ transitions, the upper states being members of a triplet electronic state having an extreme case coupling. The predissociating state A (see fig. 2) is possibly a ${}^3\Sigma$.

It is possible that the TlCl band system is due to a $1 \rightleftarrows {}^1\Sigma^+$ transition. The level which produces predissociation may be the 0^+ component of this multiplet, and the upper state A of the weak continuum at 2900 is probably analogous to the A state of TlF which is the upper state of a similar weak continuum. The ${}^1\Pi$ state in TlCl is probably the upper state of the diffuse band system, an allocation which is supported by the fact that the transition ${}^1\Pi \leftarrow {}^1\Sigma^+$ is the strongest in the spectrum. The existence of the far ultra-violet continuum is to be attributed, as with TlF, to a transition to an unstable level C.

A glance at the completed energy diagram shows that the electronic levels of TlCl are lower than the corresponding ones of TlF, which is to be expected as TlCl is the heavier molecule.

Further support for this proposed identification of the energy levels is found by examining the energy systems of the analogous molecules GaCl and InCl. The reduced masses of these molecules are of the same order, viz. GaCl 38·2, InCl 44·2 and TlCl 49·3, and therefore corresponding spectra should lie in the same region. Miescher and Wehrli (1934) find that ν_e for the $1 \rightleftarrows {}^1\Sigma^+$ transitions in the case of GaCl and InCl are 29855·7 and 28560·2 cm.$^{-1}$ respectively, these values being sufficiently close to the ν_e of the TlCl system to justify the view that this TlCl transition is also $1 \rightleftarrows {}^1\Sigma^+$ rather than ${}^1\Pi \rightleftarrows {}^1\Sigma^+$ as suggested by Miescher and Wehrli.

INTENSITY DISTRIBUTION

In Table III is shown the intensities of all the $TlCl_{35}$ heads graded from 0 to 10, the values being visual estimates made from enlargements. The most intense bands lie on a very narrow Condon parabola, a feature which is characteristic of most halide spectra.

It has been remarked that the absorption spectrum is for the most part a reversal of the emission, and as more bands have been measured in absorption than in emission the former have been chiefly used in constructing Table III. However, for all bands on the short-wave side of the 0, 0 band the intensities refer to the emission spectrum because the presence of diffuse bands in this region of the absorption spectrum produces abnormal intensities, as can be clearly seen from fig. 3 (plate 14).

TABLE III. INTENSITY DISTRIBUTION

v' \ v''	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	10	7	3												
1	2	10	8	8	3	1	1	0	0	0					
2	2	1	9	7	7	2	1	1	0	1	1				
3		1		3	3	4	2	2	1	0	1	0			
4		0	1	1	2	2	3	1	0	0	0	0	1	1	0
5					1			1	0	0	0	1	1	0	0

Such a general similarity in the emission and absorption intensities was also found with TlF, and it was then shown to be a consequence of the existence of predissociating electronic levels. It was also suggested that such a similarity in intensities could be taken as evidence for predissociation. It has already been shown that predissociation occurs with TlCl.

CLASSIFICATION

All the bands measured are classified in the usual manner in Table IV.

No data are presented for the diffuse bands as they can be found in Neujmin's paper.

The authors desire to express their gratitude to Professor W. E. Curtis, F.R.S., of King's College, Newcastle-upon-Tyne, where this work was carried out, for his interest and assistance. One of us (H. G. H.) is also greatly indebted to the Department of Scientific and Industrial Research for a grant and to Professor A. C. Menzies of University College, Southampton, for the facilities placed at his disposal for the completion of this work.

SUMMARY

The spectrum of TlCl has been studied in absorption and also in emission by means of a high-frequency discharge. Continua and diffuse bands together with a discrete band system have been found in absorption within the region 2500–3500 A. The band system also occurs in emission. It is considered that the transition is $1 \rightleftarrows {}^1\Sigma^+$, the levels having the following constants:

Level	ν	ω_e	$x_e\omega_e$
1	31054·2	216·91	6·80
${}^1\Sigma^+$	0	287·47	1·24

TABLE IV. CLASSIFICATION OF TlCl BANDS

λ (A)	I	ν	O–C	v', v''
3575·6	0	27959		4, 14 R
3566·7		28029		4, 14 R'
3557·8	0	28099		5, 14 R
3543·5	1	28213		4, 13 R
3525·2	0	28359		5, 13 R
3510·8		28475		4, 12 R
3503·8		28532		4, 12 R'
3498·26	0	28577·5		3, 11 R
3492·33	1	28626·0		5, 12 R
3487·18	1	28668·2		2, 10 R
3478·42	0	28740·5		4, 11 R
3477·40	0	28748·9		1, 9 R
3471·96		28793·9		4, 11 R'
3470·95		28802·3		1, 9 R'
3466·46	1	28839·6		3, 10 R
3461·17	1	28883·7		5, 11 R
3455·37	1	28932·1		2, 9 R
3446·80	0	29004·2		4, 10 R
3445·20	0	29017·6		1, 8 R
3441·48		29049·0		4, 10 R'
3439·96		29063·5		1, 8 R'
3434·80	0	29105·4		3, 9 R
3429·46	0	29150·7		5, 10 R and 3, 9 R'
3424·11		29196·3		5, 10 R'
3423·43	0	29202·1		2, 8 R
3415·53	0	29269·6		4, 9 R
3410·46		29313·2		4, 9 R'
3403·26	1	29375·2		3, 8 R
3401·15	0	29393·4		1, 7 R
3398·49	0	29416·4		5, 9 R and 3, 8 R'
3393·95		29456·6		5, 9 R'
3391·62	1	29476·0		2, 7 R
3387·25		29514·0		2, 7 R'
3384·56	0	29537·5		4, 8 R
3380·29		29574·8		4, 8 R'
3372·04	2	29647·2		3. 7 R
3368·38		29679·3		3, 7 R'
3367·78	0	29684·6		5, 8 R
3363·77		29720·0		5, 8 R'
3361·77		29737·7	−0·1	2, 6
3360·39	1	29749·9		2, 6 R
3357·00		29779·8		2, 6 R'
3353·93	1	29807·2		4, 7 R
3352·32		29821·5	0·8	1, 5
3350·40		29838·6		4, 7 R'
3349·98	1	29842·3		1, 5 R
3346·92		29869·6		1, 5 R'
3341·42	2	29919·2		3, 6 R
3338·32		29946·9		3, 6 R'
3337·28	1	29955·9		5, 7 R
3333·97		29985·7		5, 7 R'
3331·18		30010·8	0·8	2, 5
3329·88	2	30022·1		2, 5 R
3327·28		30045·9		2, 5 R'
3323·49	3	30080·0		4, 6 R
3322·32		30096·3	0·5	1, 4
3320·67		30105·2		4, 6 R'
3319·60	3	30115·5		1, 4 R
3317·08		30138·0		1, 4 R'
3313·53		30170·6	0·5	0, 3
3311·01		30193·6		0, 3 R and 3, 5 R
3308·70		30214·7		3, 5 R'
3300·90		30286·1	0·6	2, 4
3299·72	7	30296·8		2, 4 R
3297·77		30314·8		2, 4 R'
3293·62	2	30353·0		4, 5 R
3291·40		30374·0	0·6	1, 3
3289·61	8	30390·0		1, 3 R
3287·92		30405·6		1, 3 R'
3283·09		30450·4	0·3	0, 2
3281·40	3	30466·1		3, 4 R
3280·56	3	30473·8		0, 2 R
3279·18		30486·7		0, 2 R'
3270·92		30563·6	0·5	2, 3
3270·03	7	30572·0		2, 3 R
3268·82		30583·3		2, 3 R'
3263·60	2	30632·2		4, 4 R
3262·37		30643·8		4, 4 R'
3261·37		30652·6	−0·8	1, 2
3259·90	8	30666·9		1, 2 R
3258·94		30676·0		1, 2 R'
3253·19		30730·2	−2·4	0, 1
3251·30	3	30748·1		3, 3 R
3250·78	7	30753·0		0, 1 R
3249·76		30762·6		0, 1 R'
3248·23	1	30777·1		5, 4 R
3241·33		30842·6	−0·5	2, 2
3240·62	9	30849·4		2, 2 R
3239·99		30855·4		2, 2 R'
3233·94	1	30913·1		4, 3 R
3231·60		30935·5	−0·4	1, 1
3230·39	10	30947·1		1, 1 R
3223·15		31016·6	−1·0	0, 0
3220·87	10	31038·5		0, 0 R
3211·98		31124·4	−1·2	2, 1
3211·44	1	31129·7		2, 1 R
3205·14	1	31190·9		4, 2 R
3201·95		31222·0	1·1	1, 0
3201·62		31225·2		1, 0 R'
3201·35	2	31227·8		1, 0 R
3193·66		31303·0		3, 1 R'
3193·36	1	31305·9		3, 1 R
3183·30		31404·8		2, 0 R'
3182·64	2	31411·4		2, 0 R
3177·18		31466·8		4, 1 R'
3176·56	0	31471·5		4, 1 R

The analysis of this system given by Butkow is shown to be incorrect. Predissociation occurs in the band system and is probably responsible for the band intensities being the same in both emission and absorption. An attempt to identify the electronic levels of TlCl based upon the similarity of the spectrum to that of TlF has been made.

No sign of the Tl_{203} isotope has been detected, whereas the Cl isotope effect completely supports the proposed analysis.

REFERENCES

Butkow, K. 1929 *Z. Phys.* **58**, 232.
Howell, H. G. 1937 *Proc. Roy. Soc.* A, **160**, 242.
Jevons 1932 "Report on Band Spectra." Physical Society Report.
Miescher and Wehrli 1934 *Helv. phys. Acta*, **7**, 331.
Neujmin 1934 *Phys. Z. Sowjet.* **5**, 580.
Sponer 1936 "Molekülspektren" Part I.

β-Transitions in a coulomb field

By F. HOYLE, *Emmanuel College, Cambridge*

(*Communicated by R. H. Fowler, F.R.S.—Received* 31 *January* 1938)

1. INTRODUCTION

The theories which have been put forward (Fermi 1934; Konopinski and Uhlenbeck 1935) to explain β-decay have been worked out in detail only for allowed transitions. It is desirable to give some discussion of forbidden transitions, and to find the possible modifications of the electron energy spectra.

Following Fermi I postulate that it is possible for a neutron to transform itself into a proton, emitting also an electron and a neutrino. The whole process is supposed to take place at a given point in space. That is, the proton, electron, and neutrino all appear at the point at which the neutron disappears. To give a quantitative theory of such a process it is necessary to add to the Lagrangian of the neutron, proton, electron, and neutrino, from which their wave equations are usually derived, a new term which is linear in the wave function of each particle, and which is Lorentz invariant.

There are many terms of this type, and I have discussed previously (Hoyle 1937) the form of the electron energy spectra to which certain of them give rise. It does not seem possible,† however, to give a complete and consistent theory except when the new term in the Lagrangian contains no derivative of the wave functions (Fermi case), or perhaps, when there is a single derivative of the neutrino wave function (Konopinski-Uhlenbeck—"K.-U." case). My previous discussion (Hoyle 1937) of these latter terms containing a derivative of the neutrino wave function applies only to light nuclei in allowed transitions.

The results for heavy nuclei differ from these, in that the wave function of the electron near the nucleus is modified by the coulomb field. As Fermi has shown this dependence gives rise to a further factor

$$|\Gamma(s+1+i\gamma W(W^2-1)^{-\frac{1}{2}})/\Gamma(2s+3)|^2 e^{\pi\gamma W(W^2-1)^{-\frac{1}{2}}}(2pR/\hbar)^{2s}$$

in the transition probability, where $\gamma = z/137$ and $s = (1-\gamma^2)^{\frac{1}{2}} - 1$, p is the electron momentum, R the nuclear radius, and W is the electron energy including the rest mass in units of mc^2.

It will be shown that there are three different ways of obtaining a forbidden transition. These will be treated separately with a qualitative discussion of their relative orders of magnitude. The selection rules for the total, orbital, and spin momenta, which govern both allowed and forbidden transitions, will also be deduced. The phraseology "forbidden transition" and "allowed transition" is to be taken throughout as referring only to the order of magnitude of the decay constant. For example, "allowed transition" means "on the first Sargent curve", whilst "first and second forbidden transitions" mean "on the second and third Sargent curves" respectively.

2. THE TRANSITION PROBABILITY

Writing the wave functions of the four particles, neutron, proton, electron, and neutrino, as sv, u, ψ, ϕ respectively, the probability of β-decay in the one-body problem depends upon the matrix element

$$|\int(\psi^* A' \ldots \phi)(u^* A'' \ldots v)\,d\tau|^2,$$

where $A' \ldots$, $A'' \ldots$ are spin and derivative operators such that

$$(\psi^* A' \ldots \phi)(u^* A'' \ldots v)$$

---

† For example, the method given by Fierz (1937) cannot be applied to more than one derivative of the wave functions.

is Lorentz invariant. In the special case of a single derivative of the neutrino wave function the above matrix element may be written as

$$\left|\int\left(\psi^* A' \ldots \frac{\partial\phi}{\partial x_\mu}\right)(u^* A''_\mu \ldots v)\, d\tau\right|^2,$$

in which both $(\psi^* A' \ldots \partial\phi/\partial x_\mu$ and $(u^* A''_\mu \ldots v)$ are the $\mu \ldots$ component of a tensor. This latter expression must now be generalized for nuclear transitions, in which there may be several heavy particles capable of β-disintegration.

The initial wave function of the nucleus will be written as

$$\Phi(x_1\rho_1, x_2\rho_2, \ldots, x_m\rho_m \mid x_{m+1}\phi_{m+1}, \ldots, x_n\rho_n),$$

where x_r, ρ_r are the space and spin co-ordinates of the particle in the rth position, and the suffices $1, \ldots, m$ refer to the neutrons, and $m+1, \ldots, n$ to the protons, in the initial nucleus. The wave function of the nucleus after the neutron in the rth position has undergone a β-transition may be written as

$$\Psi(x_1\rho_1, x_2\rho_2, \ldots, x_{r-1}\rho_{r-1}, x_{r+1}\rho_{r+1}, \ldots, x_m\rho_m \mid x_r\rho_r, x_{m+1}\rho_{m+1}, \ldots, x_n\rho_n).$$

Then the probability of electron emission † will be proportional to

$$\sum_{r=1}^{m}\left|\int\left\{\sum_{\rho_1\rho_2\ldots\rho_r\rho_r'\ldots\rho_n}\Phi^*(x_1\rho_1, \ldots, x_m\rho_m \mid x_{m+1}\rho_{m+1}, \ldots, x_n\rho_n)\,(A''_\mu \ldots)_{\rho_r\rho_r'}\right.\right.$$
$$\left.\times\,\Psi(x_1\rho_1, \ldots, x_{r-1}\rho_{r-1}, x_{r+1}, \rho_{r+1}, \ldots, x_m\rho_m \mid x_r\rho_r', x_{m+1}\rho_{m+1}, \ldots, x_n\rho_n)\right\}$$
$$\left.\times\left\{\sum_{\rho\rho'}\frac{d\phi^*}{\partial x_\mu}(x_r\rho)\,(A'\ldots)_{\rho\rho'}\,\psi(x_r\rho')\right\}d\tau_1\ldots d\tau_n\right|^2,$$

the integration being carried out over the space co-ordinates $x_1, \ldots, x_n$. I wish to consider in detail chiefly the integration and summation over x_r, ρ_r, and define a new function $\chi_{\rho_r}^{\rho_r'}(x_r)$ which is equal to the above expression after omitting the summations over ρ_r, ρ_r', and the integration over x_r. This function simplifies the writing of the above expression to

$$\sum_{r=1}^{m}\left|\int\left\{\sum_{\rho_r\rho_r'}\chi_{\rho_r}^{\rho_r'}(x_r)\,(A''_\mu \ldots)_{\rho_r\rho_r'}\right\}\left\{\sum_{\rho\rho'}\frac{\partial\phi^*}{\partial x_\mu}(x_r\rho)\,(A'\ldots)_{\rho\rho'}\,\psi(x_r\rho')\right\}d\tau_r\right|^2.$$

It may be noted that since the wave functions Φ, Ψ are each antisymmetrical functions of the co-ordinates of both the neutrons and the protons, the same integral is obtained for each value of r in the above expression.

---

† This is a generalization of the usual formula in the single-body problem, and is given, for example, by Nordheim and Yost (1937).

The neutrino is a free particle, and the form of its eigen states of linear momentum are known, but before discussing the properties of the integral the positive energy solutions $\psi(x, \rho)$ of the electron wave equation are also required. A summary of the properties of such solutions is given in the Appendix, the full details having been discussed, for example, by Hulme (1931).

3. ALLOWED AND FORBIDDEN TRANSITIONS

The components of the function $\chi^{\rho_r'}_{\rho_r}(x_r)$ may be expected to have different orders of magnitude. This would arise in a similar way to the small components in the Pauli reduction of the Dirac equation for a single slow-moving particle (Bethe 1933, p. 304). In this case the $\rho = 1, 2$ components differ from the $\rho = 3, 4$ components by a factor which is of order v/c, which taking v as the velocity of a nuclear particle is about $1/10$. Now if it is assumed that a relativistic equation can be formulated for the whole nucleus, each spin co-ordinate having the values 1, 2, 3, 4, I expect a similar reduction to a non-relativistic Schrödinger equation, in which the only components of Φ, Ψ which occur will be those in which every spin co-ordinate is 3, 4. It is useful to consider the analogy with the case of a single heavy particle still further. One knows, in this latter case, that if one expresses the small components in terms of the large components, then the contribution of largest order is from the space derivatives of the large components. These space derivatives give the factor v/c. I wish to take over this result for the different components of the nuclear wave functions Φ, Ψ when the value of a given ρ is changed from 3, 4 to 1, 2 (all other spin co-ordinates having values 3, 4).

That is, if Φ', Φ'' are written for the components of Φ when the particular ρ has the values 1, 2, and 3, 4, respectively, all other ρ's having values, 3, 4, I assume that we shall then have a relation, which is, to the first order of approximation in the ratio of Φ', Φ'',

$$\Phi' = \sum_{k=1}^{3} (\sigma_k P_k) \Phi'',$$

where σ_k are the Pauli matrices, and P_k is the analogue of $\frac{1}{2}mc^2(\partial/\partial x_k + eA_k)$ in the single-body problem A_k being the vector potential (Bethe 1933, p. 305, equ. 8·4). There will be a similar relation for the components of the wave function Ψ, and if we make similar definitions this will be

$$\Psi' = \sum_{k=1}^{3} (\sigma_k Q_k) \Psi'',$$

where Q_k is similar, but not identical, to P_k. It will be of importance later to notice that the operators P_k, Q_k, may be supposed to change sign, with a change of sign of the space co-ordinates. It is now expected that if in the relativistic equations of motion for the nucleus all components of the wave function which have any spin co-ordinate except ρ_r different from 3, 4 are neglected, and if I approximate further by using a first-order expression (of the type given for Φ', Ψ') when $\rho_r = 1, 2$, then the $\rho_r = 3, 4$ components will have the same space distribution. That is, for this approximation the spin dependence upon ρ_r is independent of the space dependence upon x_r. This approximation is termed "neglecting spin forces", and this case is to be regarded as giving the first approximation. In treating the second approximation it is necessary to consider, for example, both the second-order terms in the above expressions for Φ', Ψ', and also the first-order terms in the expressions for the components of Φ, Ψ, in which two spin co-ordinates have the values 1, 2.

I shall confine myself to a discussion of the first approximation, treating first the case in which ρ_r is the spin variable of the particle making the transition. This is a special case, since the spin co-ordinate of the transition particle is operated upon by $A''_\mu \ldots$ in our matrix element. The summation, in the matrix element, over all spin co-ordinates except ρ_r, will now be only over the values 3, 4. The product function $\chi_{\rho_r}^{\rho_r'}(x_r)$ will have 16 components of which we regard 4 as "large", 8 components will differ from the "large" components by a factor of order $\frac{1}{10}$, whilst 4 components will differ by a factor of order $\frac{1}{100}$. These will be

$\chi_{\rho_r=3,4}^{\rho_r'=3,4}(x_r)$ the "large" components;

$\chi_{\rho_r=1,2}^{\rho_r'=3,4}(x_r)$ and $\chi_{\rho_r=3,4}^{\rho_r'=1,2}(x_r)$ differing by a factor of order $\frac{1}{10}$;

$\chi_{\rho_r=1,2}^{\rho_r'=1,2}(x_r)$ differing by a factor of order $\frac{1}{100}$.

Each of the functions $\chi_{\rho_r}^{\rho_r'}(x_r)$ can be expanded in a series of spherical harmonics. In the integration of these functions over the co-ordinates x_r a result which differs from zero will be obtained only from the spherical harmonic which is of the same order as the spherical harmonic given by the wave functions of the light particles. Now it is known that the light particle wave functions which have different spherical harmonic dependences have different orders of magnitude near the nucleus. The results required for these wave functions can be written down from the Appendix.

If I treat the light particle wave functions as constants, then I take G_0, F_{-1-1} for the electron, and consider the neutrino wave function as constant over the nucleus. The form of the electron energy spectrum depends only upon the form of the electron and neutrino wave functions, and this case gives a distribution which will be termed the "allowed distribution form" This is

$$A\,|\Gamma(s+1+i\gamma W(W^2-1)^{-\frac{1}{2}})/\Gamma(2s+3)\,|^2\,e^{\pi\gamma W(W^2-1)^{-\frac{1}{2}}}\,(2pR/\hbar)^{2s}$$
$$\times\,W(W^2-1)^{\frac{1}{2}}\,(W_0-W)^4.$$

The value of A will depend upon which component of $\chi^{\rho_r'}_{\rho_r}(x_r)$ is considered. Distributions of this type are obtained for all components, but A will have different orders of magnitude in accordance with the different orders of magnitude of the components of $\chi^{\rho_r'}_{\rho_r}(x_r)$.

If now only a first-order harmonic, $\cos\theta$ say, from the light particles, will give an integral different from zero, there are three ways of obtaining such a $\cos\theta$ dependence:

(1) Take F_0, G_{-1-1} for the electron, and consider the neutrino as constant over the nucleus. We shall obtain a distribution which is of the "allowed form", but the transition probability will contain a factor of order $\frac{1}{4}\gamma^2$.

(2) Take G_1, F_{-2-1} for the electron and again consider the neutrino as constant over the nucleus. The electron energy distribution will now differ from the allowed form by an extra factor (W^2-1), whilst the decay constant will be decreased by a factor of order $(\frac{1}{240})^2$.

(3) Lastly, take G_0, F_{-1-1} for the electron and consider the first-order variation of the neutrino wave function over the nucleus. This gives a distribution which differs from the allowed form by a factor $(W_0-W)^2$, and the decay constant is decreased by a factor $(\frac{1}{40})^2$ (Bethe and Bacher 1936, p. 195).

It is noted that the selection rules for these three cases will be identical, and that it is necessary to compare the orders of magnitude of the decay constants given by each case. It is convenient to write the second and third cases together as giving a distribution

$$(\tfrac{1}{40})^2\times\{\text{“allowed distribution”}\}\times\{a(W^2-1)+b(W_0-W)^2\},$$

where a, b are at most of order unity. This we call the "first modified form". It will only be comparable in magnitude with the first case when γ is very small, that is, for very light ($z \leqslant 20$) elements. This point is, however, dependent upon treating the solutions of the electron wave equation, which assumes the nucleus to act like a point charge, as a good approximation even for radii of nuclear dimensions.

There are again many possibilities if I require the possible light particle wave functions which give a second-order spherical harmonic, $\cos^2\theta$ say. These are:

(1) Take F_1, G_{-2-1}, and treat the neutrino wave function as constant.

(2) Take F_0, G_{-1-1}, and the first-order variation of the neutrino wave function over the nucleus.

(3) Take G_2, F_{-3-1}, and treat the neutrino wave function as constant.

(4) Take G_1, F_{-2-1}, and the first-order variation of the neutrino wave function over the nucleus.

(5) Take G_0, F_{-1-1}, and the second-order variation of the neutrino wave function over the nucleus.

Each of these possibilities will again have the same selection rules, and it is necessary to compare their orders of magnitude. If I take 3, 4, and 5 together I obtain approximately a distribution

$$(\tfrac{1}{40})^4 \times \{\text{“allowed distribution”}\} \times \{a(W^2-1)^2 + b(W^2-1)(W_0-W)^2 + c(W_0-W)^4\},$$

where a, b, c are at most of order unity. This distribution is termed the “second modified form”. Taking 1, 2 together gives

$$\tfrac{1}{4}\gamma^2(\tfrac{1}{40})^2 \times \{\text{“allowed distribution”}\} \times \{a(W^2-1) + b(W_0-W)^2\},$$

where a, b are again at most of order unity. It can be seen that 1, 2 give the “first modified form”, and that this will be the effective distribution except for very light elements.

The three cases where zero-, first-, and second-order spherical harmonics are taken from the wave functions of the light particles are separately considered, and it is seen that they lead not only to transition probabilities with different orders of magnitude, but also to different possible electron energy distributions. It is therefore convenient in discussing the selection rules obtained from the different components of $\chi_{\rho_r}^{\mu_r'}(x_r)$ to treat these three cases separately, since if it is supposed that for any component there is the same selection rule in two different cases (for example, in both the case where we consider the wave functions of the light particles as constants, and where they give a $\cos\theta$ angular dependence), then the only case that need be considered is that which gives the largest transition probability, that is, the case in which the light particle wave functions give the spherical harmonic of least order.

To determine the selection rules for these various transitions it is necessary to know the form of the spin operators A''_μ ..., A' ..., but before discussing the possibilities for these operators, it is useful to point out a general method for determining a selection rule. A full discussion for the case of a single particle moving in a central field is given by Dirac (1935), and also of the generalization to the many-body problem when the Hamiltonian has a vector form in the co-ordinates (Dirac 1935, § 44). In the single-body problem the magnitude of the orbital angular momentum l changes by ± 1 (in units of $\hbar$), whilst in the many-body problem we can only deduce that l changes by 0 ± 1 (if $l = 0$ then the transition $\Delta l = 0$ is forbidden). This arises because, if (L_x, L_y, L_z) is the orbital momentum vector, the expression $(L_x x + L_y y + L_z z)$ is always zero in the one-body problem but is not necessarily zero in the many-body problem.

The above work, for a Hamiltonian which is of vector form in co-ordinates (x, y, z), can be extended to the case of a Hamiltonian which contains a general vector (A_x, A_y, A_z). For if I denote by (M_x, M_y, M_z) the components of the total angular momentum i, then (M_x, M_y, M_z) will be the rotation operators for the system, and therefore the vector (A_x, A_y, A_z) will have the same commutation relations with (M_x, M_y, M_z) as (x, y, z) had with (L_x, L_y, L_z) in the case given by Dirac. It can therefore be deduced that i must change by 0 ± 1 (although if $i = 0$, $\Delta i = 0$ is forbidden). If (A_x, A_y, A_z) is a vector which depends only upon the space co-ordinates, then (L_x, L_y, L_z) will have the same commutation relations with this vector as the components of the total angular momentum, then also $\Delta l = 0 \pm 1$. If, however, (A_x, A_y, A_z) is a pure spin vector, and I define (S_x, S_y, S_z) as the components of the spin momentum S, then (S_x, S_y, S_z) will have the same commutation relations with the vector as have (M_x, M_y, M_z), and $\Delta i = 0 \pm 1$, $\Delta l = 0$, $\Delta S = 0 \pm 1$. For the cases in which (A_x, A_y, A_z) is neither a pure space nor spin vector it is necessary to resolve it, in order to determine the selection rules for l, s, into a sum of products of vectors, each of which is a pure space or spin vector. It is useful then to remember that for a non-vanishing matrix element, for a Hamiltonian which contains a vector component A_k, we must have $\Delta M_k = 0$, $\Delta L_k = 0$, $\Delta M_j = 1$, $\Delta L_j = 1$, $(j \neq k)$ if A_k depends only upon the space co-ordinates, and $\Delta M_k = 0$, $\Delta S_k = 0$, $\Delta M_j = 1$, $\Delta S_j = 1$, $(j \neq k)$ if A_k depends only upon the spin co-ordinates. The selection rules for a Hamiltonian which contains a product of vectors can be worked out by successively applying this latter result.

I. Adopting the β-decay interaction proposed by Konopinski and Uhlenbeck, we have

$$A' = \alpha_m, \quad A''_\mu = \alpha_m \gamma_\mu,$$

where α_1, α_2, α_3, α_m are the Dirac matrices, and

$$\gamma_0 = -\alpha_m, \quad \gamma_\mu = \alpha_m \alpha_\mu, \quad \mu = 1, 2, 3.$$

I have to consider the integral

$$\left| \int \left\{ \sum_{\rho_r \rho_r'} \chi^{\rho_r'}_{\rho_r}(x_r)\,(\alpha_m \gamma_\mu)_{\rho_r \rho_r'} \right\} \left\{ \sum_{\rho \rho'} \frac{\partial \phi^*}{\partial x_\mu}(x_r \rho)\,(\alpha_m)_{\rho \rho'}\,\psi(x_r \rho') \right\} d\tau_r \right|^2. \qquad (a)$$

It is convenient to consider separately the selection rules for the groups of components of $\chi^{\rho_r'}_{\rho_r}(x_r)$ with $\rho_r = 3, 4$, $\rho_r' = 3, 4$; $\rho_r = 3, 4$, $\rho_r' = 1, 2$ and $\rho_r = 1, 2$, $\rho_r' = 3, 4$; and $\rho_r = 1, 2$, $\rho_r' = 3, 4$.

(1) Treating these in turn I obtain the $\rho_r = 3, 4, \rho_r' = 3, 4$ components only with $\mu = 0$ in the above integral (a).

If I take wave functions which are constant for the light particles then I have the selection rule†

$$\Delta i = 0, \quad \Delta l = 0, \quad \Delta s = 0.$$

If now wave functions for the light particles are taken which have a first-order spherical harmonic dependence, for example, those wave functions which give $\cos\theta$, then we shall have a vector form for the Hamiltonian, and the selection rules† are

$$\Delta i = 0 \pm 1, \quad \Delta l = 0 \pm 1, \quad \Delta s = 0,$$

since $\cos\theta$ is a vector component depending only upon space co-ordinates.

The selection rules† when, for example, a $\cos^2\theta$ term is taken from the light particle wave functions, are

$$\Delta i = 0 \pm 2, \quad \Delta l = 0 \pm 2, \quad \Delta s = 0,$$

since the Hamiltonian is now a product of the components of two space vectors. It has been seen that the probability of transition in this case is much smaller than in the case where constant wave functions for the light particles are taken, and we may therefore reject here the $\Delta i = 0$, $\Delta l = 0$, $\Delta s = 0$ selection rule, since if this *is* satisfied, then the integral will not vanish in the case where we take wave functions for the light particles which are constants.

(2) The $\rho_r = 1, 2$, $\rho_r' = 1, 2$ components occur also only for $\mu = 0$ in the integral (a). If I wish only to consider elements which lie on the first, second, and third Sargent curves, then I may derive the required results by treating the light particle wave functions as constants, since these components of

---

† These selection rules are only approximate, as l, s are calculated only from the largest components of the nuclear wave functions as in the Schrödinger theory.

$\chi_{\rho_r}^{\rho_r'}(x_r)$ will already give a factor of order $(\frac{1}{100})^2$ in the decay constant. Remembering that $\chi_{\rho_r=1,2}^{\rho_r'=1,2}(x_r)$ is a product of the components of the wave functions Φ, Ψ (integration being carried out over all space co-ordinates other than x_r, and summations for the values 3, 4 over all spin co-ordinates except ρ_r), in which one spin co-ordinate only has the values 1, 2, I can express $\chi_{\rho_r=1,2}^{\rho_r'=1,2}(x_r)$ in terms of $\chi_{\rho_r=3,4}^{\rho_r'=3,4}(x_r)$ by using the relations given above for Φ'', Φ' and Ψ'', Ψ'. The Hamiltonian will now contain the factor

$$\sum_{jk=1}^{3}\left(\sum_{\rho_r\rho_r'\rho_r''=3,4}\{\chi_{\rho_r}^{\rho_r'}(x_r)\,(\sigma_k P_k)_{\rho_r\rho_r''}\,(\sigma_j Q_j)_{\rho_r''\rho'}\}\right),$$

and this factor may be written symbolically as

$$\sum_{jk=1}^{3}\{\chi(x_r)\,(\sigma_k P_k)\,(\sigma_j Q_j)\},$$

where we understand $\chi(x_r)$ to mean the function with the four components $\chi_{\rho_r=3,4}^{\rho_r'=3,4}(x_r)$. The product $(\sigma_k P_k)\,(\sigma_j Q_j)$ gives a spherically symmetrical term $(P_x Q_x + P_y Q_y + P_z Q_z)$ which gives the selection rules

$$\Delta i = 0, \quad \Delta l = 0, \quad \Delta s = 0,$$

and also terms of the type $\sigma_z(P_x Q_y - P_y Q_x)$ which is the product of the z components of a spin vector, and an axial space vector. These latter terms give the selection rules,

$$\Delta i = 0, \quad \Delta l = 0 \pm 1, \quad \Delta s = 0 \pm 1.$$

Finally, I have to examine the selection rules for the components $\chi_{\rho_r=1,2}^{\rho_r'=3,4}(x_r)$ and $\chi_{\rho_r=3,4}^{\rho_r'=1,2}(x_r)$. I must now put $\mu = 1, 2, 3$ in (A_μ), and use either the relation between Φ'', Φ' or Ψ'', Ψ', obtaining Hamiltonian terms which can be written as

$$\sum_{k=1}^{3}\chi(x_r)\,(\sigma_k A_k)\,\sigma_j,$$

where $\chi(x_r)$ again represents the four components $\chi_{\rho_r=3,4}^{\rho_r'=3,4}(x_r)$, and A_k is one or other of the vectors P_k, Q_k. These components will contribute a factor $\frac{1}{100}$ to the transition probability, and it is necessary only to consider the two cases where the light particle wave functions give a constant term, and a first-order spherical harmonic, in order to include all elements on Sargent curves not lower than the third. If the light particle wave functions are treated as constants, then our selection rules are

$$\Delta i = 0 \pm 1, \quad \Delta l = 0 \pm 1, \quad \Delta s = 0 \pm 1,$$

but if we take, for example, a $\cos\theta$ dependence, then the selection rules will be

$$\Delta i = 0 \pm 1 \pm 2, \quad \Delta l = 0 \pm 1 \pm 2, \quad \Delta s = 0 \pm 1,$$

and if $i = 0$, $\Delta i = 0$ is forbidden.

I now collect the results, on the form of the electron energy distributions, the order of magnitude of the decay constant, and the selection rules. The odd-even character of the transitions has been omitted above, but is easily added when we remember that $\cos\theta$ is odd, $\cos^2\theta$ is even, and P_k, Q_k are both odd, all for changes of sign of the space co-ordinates.

It is now easy to see from Table I to which forbidden transition the various selection rules correspond. Many of the selection rules can be discarded as having no practical application, since for a given forbidden transition it is necessary only to consider the selection rules which differ from those for forbidden transitions of lower order.

TABLE I. THE SELECTION RULES AND ELECTRON DISTRIBUTIONS IN THE K.-U. CASE

Components of $\chi_{\rho_r}^{\rho_r'}(x_r)$ considered	Taking light particle wave functions which are constants	Taking a $\cos\theta$ dependence from the light particle wave functions	Taking a $\cos^2\theta$ dependence from the light particle wave functions
$\chi_{\rho_r=3,4}^{\rho_r'=3,4}(x_r)$	$\Delta i=0$, $\Delta l=0$, $\Delta s=0$ Even-even Odd-odd $(a.d.f.)$	$\Delta i=0\pm 1$, $\Delta l=0\pm 1$, $\Delta s=0$ Odd-even Even-odd $\frac{1}{4}\gamma^2 \times (a.d.f.)$ and $(\frac{1}{40})^2\,(m_1f)$	$\Delta i=0\pm 2$, $\Delta l=0\pm 2$, $\Delta s=0$ Even-even Odd-odd $\frac{1}{4}\gamma^2\,(\frac{1}{40})^2\,(m_1f)$ and $(\frac{1}{40})^4\,(m_2f)$
$\chi_{\rho_r=3,4}^{\rho_r'=1,2}(x_r)$ $\chi_{\rho_r=1,2}^{\rho_r'=3,4}(x_r)$	$\Delta i=0\pm 1$, $\Delta l=0\pm 1$, $\Delta s=0\pm 1$ Odd-even Even-odd $(\frac{1}{100})\cdot(a.d.f.)$	$\Delta i=0\pm 1\pm 2$, $\Delta l=0\pm 1\pm 2$, $\Delta s=0\pm 1$ Even-even Odd-odd $\frac{1}{4}\gamma^2(\frac{1}{100})\,(a.d.f.)$ and $(\frac{1}{100})\,(\frac{1}{40})^2\,(m_1f)$	
$\chi_{\rho_r=1,2}^{\rho_r'=1,2}(x_r)$	$\Delta i=0$, $\Delta l=0\pm 1$, $\Delta s=0\pm 1$ Even-even Odd-odd $(\frac{1}{100})^2\,(a.d.f.)$		If in any transition $i=0$, $\Delta i=0$ is forbidden

$a.d.f.$ = "allowed distribution form".
m_1f = "first modified form".
m_2f = "second modified form".

It is seen that there are different components of $\chi^{\rho_r'}_{\rho_r}(x_r)$ which will give the same forbidden transition although the components themselves have different orders of magnitude. A comparison of the actual orders of magnitude of the transition probabilities for these different components will only occur when the transitions have the same selection rules. This is of importance when the possible forms of the electron energy spectra are required. Remembering this, it is seen from the table that it is not necessary for either heavy or light elements on the second or third Sargent curves to have electron spectra which differ effectively from the "allowed distribution form".

There are now other possibilities for the operators $A' \ldots,\ A''_\mu \ldots$. The discussion of these will be very similar to that of the K.-U. case, and I shall treat these various possibilities in less detail.

II. With $\quad A''_\mu = \alpha_m \gamma_\mu \gamma_5, \quad A' = \alpha_m \gamma_5, \quad \gamma_5 = \alpha_1 \alpha_2 \alpha_3,$

I obtain contributions from the $\chi^{\rho_r'=3,4}_{\rho_r=3,4}(x_r)$ components when $\mu \neq 0$, and for these components of $\chi^{\rho_r'}_{\rho_r}(x_r)$ if the light particle wave functions are treated as constants, the selection rules are

$$\Delta i = 0 \pm 1, \quad \Delta l = 0, \quad \Delta s = 0 \pm 1.$$

With a first-order spherical harmonic, $\cos\theta$ say, from the light particle wave functions, the selection rules are

$$\Delta i = 0 \pm 1 \pm 2, \quad \Delta l = 0 \pm 1, \quad \Delta s = 0 \pm 1,$$

and again, with $\cos^2\theta$ from the wave functions, the selection rules are

$$\Delta i = 0 \pm 1 \pm 2 \pm 3, \quad \Delta l = 0 \pm 1 \pm 2, \quad \Delta s = 0 \pm 1.$$

If now I consider the $\chi^{\rho_r'=1,2}_{\rho_r=1,2}(x_r)$ components of $\chi^{\rho_r'}_{\rho_r}(x_r)$, then, as before, the vectors P_k, Q_k are introduced, and with the light particle wave functions constant, I have

$$\Delta i = 0 \pm 1, \quad \Delta l = 0 \pm 1 \pm 2, \quad \Delta s = 0 \pm 1.$$

Finally, for the $\chi^{\rho_r'=3,4}_{\rho_r=1,2}(x_r)$ and $\chi^{\rho_r'=1,2}_{\rho_r=3,4}(x_r)$ components put $\mu = 0$ in (A''_μ) and introduce, as in the K.-U. case, one or other of the vectors P_k, Q_k. With constant wave functions for the light particles, the selection rules are

$$\Delta i = 0, \quad \Delta l = 0 \pm 1, \quad \Delta s = 0 \pm 1.$$

If, however, a $\cos\theta$ dependence from the light particle is taken the selection rules become

$$\Delta i = 0 \pm 1, \quad \Delta l = 0 \pm 1 \pm 2, \quad \Delta s = 0 \pm 1,$$

and again a table is constructed for the results.

TABLE II. THE SELECTION RULES AND ELECTRON DISTRIBUTIONS FOR THE INTERACTION II

Components of $\chi_{\rho_r}^{\rho_r'}$ considered	Taking light particle wave functions which are treated as constants over the nucleus	Taking a cos θ dependence from the light particle wave functions	Taking a cos² θ dependence from the light particle wave functions
$\chi_{\rho_r = 3,4}^{\rho_r' = 3,4}(x_r)$	$\Delta i = 0 \pm 1$, $\Delta l = 0$, $\Delta s = 0 \pm 1$ Even-even Odd-odd $(a.d.f.)$	$\Delta i = 0 \pm 1 \pm 2$, $\Delta l = 0 \pm 1$, $\Delta s = 0 \pm 1$ Even-odd Odd-even $\frac{1}{4}\gamma^2(a.d.f.)$ and $\left(\frac{1}{40}\right)^2 (m_1 f)$	$\Delta i = 0 \pm 1 \pm 2 \pm 3$, $\Delta l = 0 \pm 1 \pm 2$, $\Delta s = 0 \pm 1$ Even-even Odd-odd $\frac{1}{4}\left(\frac{1}{40}\right)^2 \gamma^2(m_1 f)$ and $\left(\frac{1}{40}\right)^4 (m_2 f)$
$\chi_{\rho_r = 3,4}^{\rho_r' = 1,2}(x_r)$ $\chi_{\rho_r = 1,2}^{\rho_r' = 3,4}(x_r)$	$\Delta i = 0$, $\Delta l = 0 \pm 1$, $\Delta s = 0 \pm 1$ Odd-even Even-odd $\left(\frac{1}{100}\right)(a.d.f.)$	$\Delta i = 0 \pm 1$, $\Delta l = 0 \pm 1 \pm 2$, $\Delta s = 0 \pm 1$ Even-even Odd-odd $\frac{1}{4}\left(\frac{1}{100}\right)\gamma^2(a.d.f.)$ and $\frac{1}{100}\left(\frac{1}{40}\right)^2$ $(m_1 f)$	
$\chi_{\rho_r = 1,2}^{\rho_r' = 1,2}(x_r)$	$\Delta i = 0 \pm 1$, $\Delta l = 0 \pm 1 \pm 2$, $\Delta s = 0 \pm 1$ Even-even Odd-odd $\left(\frac{1}{100}\right)^2 (a.d.f.)$		If $i = 0$, $\Delta i = 0$ is forbidden

$a.d.f.$ = "allowed distribution form".
$m_1 f$ = first modified form.
$m_2 f$ = second modified form.

It is again seen from Table II that it is possible to have a forbidden transition for which the electron energy distribution does not differ effectively from the "allowed form". If transitions are considered which satisfy, for example, the selection rules $\Delta i = 1$, $\Delta l = 2$, $\Delta s = 1$, and which are even-even or odd-odd in character, then we must have a second forbidden transition. The chief contribution to the electron energy spectrum will be from the $\chi_{\rho_r = 1,2}^{\rho_r' = 1,2}(x_r)$ components, since the decay constant for these components receives a factor of order $\left(\frac{1}{100}\right)^2$, whilst the decay constant for transitions from all other components must receive, both for heavy and light elements, a factor which is small compared with $\left(\frac{1}{100}\right)^2$. Thus the electron energy distribution will have effectively the "allowed form", but with a slight humping, caused by the contributions from the other components, at the upper and lower energy limits. If, on the other hand, I suppose the transitions to satisfy the selection rules $\Delta i = 3$, $\Delta l = 2$, $\Delta s = 1$, and to have an even-even or odd-odd character, then it is seen that the corresponding electron spectrum will differ widely from the "allowed form".

III. I now discuss a third possibility for $A''_{\mu} \ldots$ and $A' \ldots$,

$$A''_{\mu\nu} = \alpha_m \gamma_\mu \gamma_\nu, \quad A'_\nu = \alpha_m \gamma_\nu.$$

The $\chi_{\rho_r=3,4}^{\rho_r'=3,4}$ components are obtained when μ, $\nu = 1, 2, 3$ but $\mu \neq \nu$. These terms will have the same vector character as the terms which gave the $\chi_{\rho_r=3,4}^{\rho_r'=3,4}(x_r)$ components in the case II considered above, and we may thus take over all selection rules from Table I for both $\chi_{\rho_r=3,4}^{\rho_r'=3,4}(x_r)$ and $\chi_{\rho_r=1,2}^{\rho_r'=1,2}(x_r)$.

TABLE III. THE SELECTION RULES AND THE ELECTRON DISTRIBUTIONS FOR THE INTERACTION III

Components of $\chi_{\rho_r}^{\rho_r'}(x_r)$ considered	Taking light particle wave functions which are treated as constants over the nucleus	Taking a $\cos\theta$ dependence from the light particle wave functions	Taking a $\cos^2\theta$ dependence from the light particle wave functions
$\chi_{\rho_r=3,4}^{\rho_r'=3,4}(x_r)$	$\Delta i = 0 \pm 1$, $\Delta l = 0$, $\Delta s = 0 \pm 1$ Even-even, Odd-odd (*a.d.f.*)	$\Delta i = 0 \pm 1 \pm 2$, $\Delta l = 0 \pm 1$, $\Delta s = 0 \pm 1$ Even-odd, Odd-even $\frac{1}{4}\gamma^2$(*a.d.f.*) and $(\frac{1}{40})^2$ ($m_1 f$)	$\Delta i = 0 \pm 1 \pm 2 \pm 3$, $\Delta l = 0 \pm 1 \pm 2$, $\Delta s = 0 \pm 1$ Even-even, Odd-odd $\frac{1}{4}(\frac{1}{40})^2\gamma^2$($m_1 f$) and $(\frac{1}{40})^4$ ($m_2 f$)
$\chi_{\rho_r=1,2}^{\rho_r'=3,4}(x_r)$ $\chi_{\rho_r=3,4}^{\rho_r'=1,2}(x_r)$	$\Delta i = 0 \pm 1$, $\Delta l = 0 \pm 1$, $\Delta s = 0 \pm 1$ Odd-even, Even-odd $(\frac{1}{100})$ (*a.d.f.*)	$\Delta i = 0 \pm 1 \pm 2$, $\Delta l = 0 \pm 1 \pm 2$, $\Delta s = 0 \pm 1$ Even-even, Odd-odd $\frac{1}{4}(\frac{1}{100})\gamma^2$(*a.d.f.*) and $(\frac{1}{100})(\frac{1}{40})^2$ ($m_1 f$)	
$\chi_{\rho_r=1,2}^{\rho_r'=1,2}(x_r)$	$\Delta i = 0 \pm 1$, $\Delta l = 0 \pm 1 \pm 2$, $\Delta s = 0 \pm 1$ Even-even, Odd-odd $(\frac{1}{100})^2$ (*a.d.f.*)		

a.d.f. = "allowed distribution form".
$m_1 f$ = "first modified form".
$m_2 f$ = "second modified form".

In considering $\chi_{\rho_r=1,2}^{\rho_r'=3,4}(x_r)$ and $\chi_{\rho_r=3,4}^{\rho_r'=1,2}(x_r)$ put $\mu = 0$ or $\nu = 0$, and these terms will again have the same vector character as the corresponding terms in the K.-U. case. Therefore the K.-U. selection rules for the components $\chi^{\rho_r'=3,4}_{\rho_r=1,2}(x_r)$ and $\chi^{\rho_r'=1,2}_{\rho_r=3,4}(x_r)$ may be taken over, and the table for this choice of $A''_\mu \ldots$, $A' \ldots$, is obtained immediately from our two previous tables.

IV. Lastly, we may have

$$A''_{\mu\nu} = \alpha_m \gamma_\mu \gamma_\nu \gamma_5, \quad A'_\nu = \alpha_m \gamma_\nu \gamma_5.$$

The $\chi^{\rho_r'=3,4}_{\rho_r=3,4}(x_r)$ and $\chi^{\rho_r'=1,2}_{\rho_r=1,2}(x_r)$ components are given by either $\mu = 0$ or $\nu = 0$, and the same vector character for these components is obtained as in case II. Therefore the appropriate selection rules may be taken over from Table II.

To obtain the $\chi^{\rho_r'=3,4}_{\rho_r=1,2}(x_r)$ and $\chi^{\rho_r'=1,2}_{\rho_r=3,4}(x_r)$ components I must have μ, ν both different from zero, and once more this gives the same vector character as in the K.-U. case. Thus the table for this form of $A''_\mu \ldots$, $A' \ldots$ will be exactly of the same form as Table III.

In the above discussion of the possibilities for $A''_\mu \ldots$, $A' \ldots$ I have neglected all components of the wave functions Φ, Ψ which have any spin co-ordinate except ρ_r with the values 1, 2, where ρ_r is the spin co-ordinate of the heavy particle making the transition. This means that the whole change of momentum comes from the transition particle. In the same approximation I must, however, consider the case in which the spin co-ordinate which takes the values 1, 2 is not the spin co-ordinate of the transition particle. The only components of $\chi^{\rho_r'}_{\rho_r}(x_r)$ which occur in this case are those for which $\rho_r = 3, 4$, $\rho_r' = 3, 4$, but we remember that $\chi^{\rho_r'}_{\rho_r}(x_r)$ is the sum of products of the components of the wave functions Φ, Ψ where these are summations over all spin co-ordinates other than ρ_r, ρ_r', and integrations over all space co-ordinates other than x_r. The summations will now all be over the values 3, 4 with the exception of the particular spin co-ordinate to which all of the values 1, 2, 3, 4 are allowed, and this special summation will be of exactly the same type as the spin summation considered above for the K.-U. term with $\mu = 0$ in the interaction $A''_\mu \ldots$. Thus the various components in this summation will contribute the selection rules given in the first and third rows of the first column of Table I (there being no selection rules similar to the second row, since no suitable components arise in the summation, and none similar to the second and third columns, since these are due to the properties of the light particle wave functions). It is then seen that the total selection rules for the transition will depend upon the properties of this special summation, and it will depend also upon the summations over the spin co-ordinates of the transition particles, that is, over ρ_r, $\rho_r' = 3, 4$. There will be a contribution from these latter summations both on account of the interaction operator A''_μ ... and also the different possible space dependences of light particle wave functions. In general all possible total selection rules for these transitions are obtained by combining the first row of the table corresponding to the interaction form chosen with the first and third rows of the first column of Table I.

Thus if the interaction II is chosen, and light particle wave functions which can be treated as constant over the nucleus are taken, we have the possibilities

$$\Delta i = 0 \pm 1, \quad \Delta l = 0, \quad \Delta s = 0 \pm 1,$$

giving an "allowed distribution form" for the electron spectrum, and also

$$\Delta i = 0 \pm 1, \quad \Delta l = 0 \pm 1, \quad \Delta s = 0 \pm 1 \pm 2,$$

again giving an electron spectrum with an "allowed form" but now giving a factor of order $(\frac{1}{100})^2$ in the decay constant. It may be noted that the factor in the decay constant will be given by the product of the separate factors in the two sections of the tables which are combined, whilst the form of the electron energy spectrum is given solely by the section which corresponds to the co-ordinates of the transition particles. The case $\Delta s = \pm 2$ gives a new possibility not realized in any of the above tables. For the former transitions whose selection rules were given by the separate tables, we had the result that, for first and second forbidden transitions, the electron energy spectrum may still have effectively the "allowed distribution form". We have now a similar result for these latter transitions, and also since the selection rule $\Delta s = \pm 2$ can be obtained only in the case given above (provided the properties of elements only down to the third Sargent curve are required, and the interaction II is assumed), it is seen that the electron spectrum has exactly the "allowed form", that is, there are no slight divergences from the "allowed form" at the upper and lower energy limits. This conclusion does not hold, however, for a K.-U. interaction, but will hold for any of the other possibilities.

I have now discussed the problem of first and second forbidden transitions in first approximation, but to the order of the second approximation in the "spin force terms", the selection rules deduced for the various possible interaction forms, may be broken. If, however, the spin force terms are small, that is, if the approximation considered is a good approximation, then it may be expected that the arguments on the possible forms of the electron spectra will hold. That is, we should expect a number of elements on both the second and third Sargent curves to give electron distributions of the "allowed form".

My warmest thanks are due to Professor Peierls, both for many suggestions upon this problem, and for all I have learned from him. I am also greatly indebted to the Goldsmiths Company for their financial assistance.

APPENDIX

The electron wave function

The electron can be considered as moving in a central field with potential $v = ze/r$. The solutions of Dirac's equation in this case can be written in the two forms

$$\Psi_1 = -iF_k P^u_{k+1}, \qquad \Psi_2 = -iF_k P^{u+1}_{k+1},$$

$$\Psi_3 = (k+u+1)\, G_k P^u_k, \qquad \Psi_4 = (-k+u)\, G_k P^{u+1}_k,$$

and

$$\Psi_1 = -i(k+u)\, F_{-k-1} P^u_{k-1}, \qquad \Psi_2 = -i(-k+u+1)\, F_{-k-1} P^{u+1}_{k-1},$$

$$\Psi_3 = G_{-k-1} P^u_k, \qquad \Psi_4 = G_{-k-1} P^{u+1}_k,$$

where F_k, G_k satisfy

$$\frac{1}{\hbar}\left(\frac{W+eV}{c}+mc\right)F_k + \frac{dG_k}{dr} - \frac{k}{r}G_k = 0,$$

$$-\frac{1}{\hbar}\left(\frac{W+eV}{c}-mc\right)G_k + \frac{dF_k}{dr} + \frac{k+2}{r}F_k = 0,$$

with $+k \geqslant 0$, $-k < 0$, and

$$P^u_k = (k-u)\,!\,\sin^u\theta\left(\frac{d}{d\cos\theta}\right)^{k+u}\frac{(\cos^2 u-1)^k}{2^k k\,!}\,e^{iu\phi},$$

u, k being positive integers such that $k \geqslant u$. The above solutions have energy W and angular momentum $u+\frac{1}{2}$, the quantum numbers being W, k and u.

If I introduce the definitions

$$\mathscr{F}_k = AF_k - BG_k,$$

$$\mathscr{G}_k = AF_k + BG_k,$$

where

$$A^2 = 1/\hbar(mc + W/c),$$

$$B^2 = 1/\hbar(mc - W/c),$$

then for small distances from the origin, for example of the order of the nuclear radius, the solutions of the above equations for F_k, G_k are

$$\mathscr{G}_k = -2^{2s+2}[(-k+s)+i(b+c)]$$
$$(s-ib+1)\,Nr^s a^{2s+2}\,e^{-\pi b}\,|\,\Gamma(s-ib+1)\,|^2/\Gamma(2s+3),$$

$$\mathscr{F}_k = 2^{2s+2}[(k-s)+i(b+c)]$$
$$(s+ib+1)\,Nr^s a^{2s+2}\,e^{-\pi b}\,|\,\Gamma(s-ib+1)\,|^2/\Gamma(2s+3),$$

$$\mathscr{F}_{-k-1} = 2^{2s'+2}\,[(b-c)-i(k-1-s')]$$
$$(s'+ib+1)\,N'r^{s'}a^{2s'+2}\,e^{-\pi b}\,|\,\Gamma(s'-ib+1)\,|^2/\Gamma(2s'+3),$$

$$\mathscr{G}_{-k-1} = 2^{2s'+2}\,[(b-c)+i(k-1-s')]$$
$$(s'-ib+1)\,N'r^{s'}a^{2s'+2}\,e^{-\pi b}\,|\,\Gamma(s'-ib+1)\,|^2/\Gamma(2s'+3),$$

where N, N' are suitable factors which normalize these solutions with respect to the energy, and

$$s = \{(k+1)^2-\gamma^2\}^{\frac{1}{2}}-1, \quad s' = \{k^2-\gamma^2\}^{\frac{1}{2}}-1, \quad \gamma = z/137,$$
$$b = \gamma/2(A^2+|\,B\,|^2)/A\,|\,B\,|, \quad c = \gamma/2(A^2-|\,B\,|^2)/A\,|\,B\,|, \quad a = A\,|\,B\,|.$$

The normalization factor N is given by

$$N = (W/\hbar c^2 a)^{\frac{1}{2}}\,A\,|\,B\,|/(A^2+|\,B\,|^2)^{\frac{1}{2}}\,.\,K^{-1},$$

where

$$K = 1/2\{(k-s)^2+(b+c)^2\}^{\frac{1}{2}}\,|\,\Gamma(s+ib+1)\,|\,e^{-3\pi b/2}(2a)^{s+1}\{4\pi(k+u+1)!\,(k-u)!\}^{\frac{1}{2}}$$

and N' is given by replacing k by $k-1$, s by s', and c by $-c$. We have

$$F_k = 1/2A(\mathscr{F}_k+\mathscr{G}_k), \quad G_k = i/2\,|\,B\,|(\mathscr{F}_k-\mathscr{G}_k),$$

giving

$$F_k = 2^{2s+2}/A[(k-s)\,(s+1)-b(b+c)]$$
$$Nr^s a^{2s+2}\,e^{-\pi b}\,|\,\Gamma(s-ib+1)\,|^2/\Gamma(2s+3),$$

$$G_k = -2^{2s+2}/|\,B\,|\,[(k-s)\,b+(b+c)\,(s+1)]$$
$$Nr^s a^{2s+2}\,e^{-\pi b}\,|\,\Gamma(s-ib+1)|^2/\Gamma(2s+3),$$

$$F_{-k-1} = 2^{2s'+2}/A[(b-c)\,(s'+1)+(k-1-s')\,b]$$
$$N'r^{s'}a^{2s'+2}\,e^{-\pi b}\,|\,\Gamma(s'-ib+1)\,|^2/\Gamma(2s'+3),$$

$$G_{-k-1} = -2^{2s'+2}/|\,B\,|\,[(b-c)\,b-(k-1-s')\,(s'+1)]$$
$$N'r^{s'}a^{2s'+2}\,e^{-\pi b}\,|\,\Gamma(s'-ib+1)\,|^2/\Gamma(2s'+3).$$

If γ^4 and higher powers are neglected, then

$$s = k - \gamma^2/2(k+1), \text{ that is } (k-s)(s+1) = \gamma^2/2,$$

and $$s' = k(1-\gamma^2/2k^2) - 1, \text{ that is } (k-1-s')(s'+1) = \gamma^2/2,$$

and since $$(b+c) = \gamma A/|B|, \quad (b-c) = \gamma |B|/A,$$

$$\begin{aligned}(k-s)(s+1) - b(b+c) &= \gamma^2/2 - \gamma^2/2(A/|B|)(A^2+|B|^2)/(A|B|) \\ &= -\tfrac{1}{2}\gamma^2(A^2/|B|^2), \\ (b-c)b - (k-1-s')(s'+1) &= \gamma^2/2(|B|/A)(A^2-|B|^2)/(A|B|) - \gamma^2/2 \\ &= -\tfrac{1}{2}\gamma^2(|B|^2/A^2).\end{aligned}$$

Thus F_k differs from G_k by a factor of order $\dfrac{1}{2(s+1)}\cdot\gamma\left(\dfrac{A}{|B|}\right)$ and G_{-k-1} differs from F_{-k-1} by a factor of order $\dfrac{1}{2(s'+1)}\cdot\gamma\left(\dfrac{|B|}{A}\right)$.

In evaluating transition probabilities one can neglect the contribution from all electron wave functions other than G_0 and F_{-1-1}, unless those integrals to which G_0, F_{-1-1} contribute happen to vanish. For example, this occurs when the other factors in the integral, averaged over direction, vanish. In that case only the electron wave functions give a contribution which contain a factor $\cos\theta$, or $\cos^2\theta$, etc., as the case may be. For heavy elements ($z \geqslant 20$), the largest wave functions containing $\cos\theta$ are F_0, G_{-1-1} and these are of order $\gamma^2/4$ (for very light elements the largest wave functions will be G_1, F_{-1-1}), whilst terms containing $\cos^2\theta$ arise only for higher values of k. G_1, F_{-2-1} differ from G_0, F_{-1-1} by a factor which is of order

$$2(2ra)^{1+\gamma^2/4}\,|\,\Gamma(2-ib)\,\Gamma(3)/\Gamma(5)\,\Gamma(-ib+1)\,|,$$

(G_k has the dependence upon k given by the factor

$$(2ra)^s(s+1)\,|\,\Gamma(s-ib+1)\,|/\Gamma(2s+3)\{(k+u+1)!\,(k-u)!\}^{\frac{1}{2}},$$

whilst F_{-k-1} contains a similar factor), which if $r \sim 9\times10^{-13}$ cm. gives approximately the value

$$\tfrac{1}{240}(W^2-1)^{\frac{1}{2}} \text{ for } u=0.$$

The relative orders of magnitude of the electron wave-functions which we require down to the second forbidden transition have now been obtained. It may be noted that

$$F^2_{-1-1} + G^2_0 = 2^{4s+4}(1/A^2 + 1/|B|^2)\,N^2 r^{2s} a^{4s+4} e^{-2\pi b}\{|\,\Gamma(s-ib+1)\,|^2/\Gamma(2s+3)\}^2$$

with $s = (1+\gamma^2)^{\frac{1}{2}} - 1$.

Substituting for N^2 gives, apart from a numerical factor,

$$|\Gamma(s+ib+1)/\Gamma(2s+3)|^2\, W(W^2-1)^{\frac{1}{2}}\, e^{\pi W(W^2-1)^{-\frac{1}{2}}}\, (2rp/\hbar)^{2s},$$

since $$a = A/|B| = (W^2/c^2 - m^2c^2)^{\frac{1}{2}}\, 1/\hbar = (W^2-1)^{\frac{1}{2}}\, 1/\hbar c = p/\hbar$$

and $$b = \gamma W(W^2-1)^{-\frac{1}{2}}.$$

This is the factor mentioned above in the introductory remarks. $(F_0^2 + G^2_{-1-1})$ similarly contains this factor with the additional $\gamma^2/4$.

It may be pointed out, finally, that if the wave functions for $k = 0$ and $k = 1$ are denoted by ψ_0 and ψ_1, then for small r

$$|\psi_1| \ll |\psi_0|,$$

but that $$\left|\frac{\partial\psi_0}{\partial r}\right| \ll \left|\frac{\partial\psi_1}{\partial r}\right|.$$

SUMMARY

This paper attempts to give the selection rules, and the possible forms of the electron energy spectra, which correspond to elements on the first, second, and third Sargent curves, in the case of each of the possible forms of β-interaction belonging to Hamiltonians that contain a derivative of only the neutrino wave function. This allows a choice of several possibilities for the interaction, among which is the form proposed by Konopinski and Uhlenbeck. The accurate solution of the problem would require a knowledge of the wave equation of a nucleus containing many particles. I assume that a non-relativistic Schrödinger equation can be formulated for the nucleus, in which the spin co-ordinate of each particle has two possible values (3, 4). The solutions of such equations for the initial and final nuclei will give a first type of forbidden transition. It is further assumed that a relativistic equation can be constructed for the nuclei, in which the spin co-ordinate of each particle will now have four values (1, 2, 3, 4). From this point of view I regard the Schrödinger equations as given by neglecting, in the relativistic equations, all components of the wave function in which any spin co-ordinate is 1, 2. The Pauli reduction of the Dirac equation in the single-body problem is generalized to the assumption that a component of the wave function in which one spin co-ordinate is different from 3, 4 can be expressed to a suitable approximation, in terms of those components which occur in the Schrödinger equation, the connexion between these components being analogous to that given by Pauli in the one-body problem. This introduces small components of the wave functions into the expression for the transition

probability. These small components will have selection rules which are different from those for the large components, and transitions which were forbidden may now become "allowed" for the small components (that is, the light particle wave functions may be treated as constants over the nucleus). It is convenient to distinguish those components which are small in the spin variable of the transition particle, and those which are small in other variables. These two groups of small components will, in general, also have different selection rules. The result of comparing the selection rules for the three groups of components of the wave functions (the large components, and the two groups of small components), and of discussing the corresponding forms of the electron spectra, show that there may be elements on the second Sargent curve with either

(1) The "allowed distribution form" given by Konopinski and Uhlenbeck, or possibly for light elements (nuclear charge $\leqslant 20$):

(2) Electron distributions which differ from (1), and of types previously discussed (Hoyle 1937, p. 290, fig. 4, I, II),

and that for elements on the third Sargent curve we have the possibilities:

(1) The "allowed distribution form" given by Konopinski and Uhlenbeck.

(2) Effectively the distribution (1), but with a slight humping at the upper and lower energy limits.

(3) Distributions of the types previously discussed (Hoyle 1937, fig. 4, I, II).

(4) And for light elements (nuclear charge $\leqslant 20$) distributions which differ more widely from the "allowed form" than (3), the shapes of these distributions being similar to (3), but of a more exaggerated form.

REFERENCES

Bethe, H. A. 1933 *Handbuch der Physik*, **24**/I, 273–560.
Bethe, H. A. and Bacher, R. F. 1936 *Rev. Mod. Phys.* **8**, 83–229.
Dirac, P. A. M. 1935 "Principles of Quantum Mechanics."
Fermi, E. 1934 *Z. Phys.* **88**, 161–77.
Fierz, M. 1937 *Helv. phys. Acta*, **3**.
Hoyle, F. 1937 *Proc. Camb. Phil. Soc.* **33**, 277–92.
Hulme, H. R. 1931 *Proc. Roy. Soc.* A, **133**, 381–406.
Konopinski, E. J. and Uhlenbeck, G. E. 1935 *Phys. Rev.* **48**, 7–12.
Nordheim, L. W. and Yost, F. L. 1937 *Phys. Rev.* **51**, 943–7.

The electrical conductivity of thin metallic films
III. Alkali films with the properties of the normal metal

BY A. C. B. LOVELL

Physical Laboratories, University of Manchester

(*Communicated by A. M. Tyndall, F.R.S.—Received 4 February* 1938)

1. INTRODUCTION

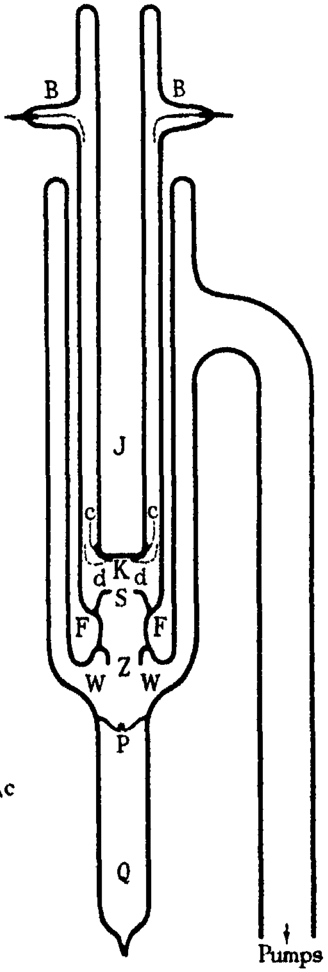
In the present work the measurements of the resistivity of evaporated films of the alkali metals (Lovell 1936*a*, *b*; Appleyard and Lovell 1937) have been extended to a thickness of several thousand angstroms. The previous results were limited to films of less than a few hundred angstroms in thickness; and it appeared that the increased resistivity of these thin films compared with that of the bulk metal was due merely to the shortening of the mean free path of the conduction electrons by collision with the boundaries of the film. It was therefore to be expected that at thicknesses much in excess of the mean free path the resistivity of the film material should closely approximate to that of the bulk metal.

The present results confirm this expectation only in the case of caesium, where the resistivity approaches within a few per cent that of the bulk metal. Potassium and rubidium films, on the other hand, show a considerable excess of resistivity above that of the bulk metal, but it is shown that this excess is a residual resistance due to lattice strain, and that it may be partially removed by suitable annealing.

2. EXPERIMENTAL

The modified experimental tube is shown in fig. 1. Comparison with fig. 1 (Lovell 1936*b*) shows that its essential features are the same as those of the previous tube, though there are considerable alterations of detail.

The cooled surface on which condensation takes place is the bottom of the long re-entrant pocket *J*, and the cooling bath *FF* has been extended upwards so as to enclose a considerable length of *J*. This construction allows the condensing surface to be brought closer to the pinhole *P*, and so permits of a much higher rate of deposition without raising the vapour pressure in the oven. In this way it was possible to deposit films of many thousand angstroms in thickness within a comparatively short period, and without the risk of the alkali metal attacking the pyrex-glass oven through raising the temperature. Further, the increased height of the cooling bath *FF* improves the heat insulation of the liquid nitrogen in *J*, and permits of long depositions without replenishing; a convenience when a liquefied gas is being boiled under reduced pressure.



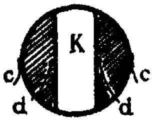


FIG. 1a

FIG. 1. The current leads *cc* and potential leads *dd* are brought out by four pinches around the periphery of the tube at *BB*.

The defining slit *S* was narrowed to 2·1 mm., so that the condensate now formed a narrow strip (2·6 × 11·4 mm.) overlapping the graphited contacts.

The effects of distortions of current flow where the current entered and left the film (due to possible unevenness of the graphite coating) were thus minimized. An additional subsidiary slit Z helped to cut down scattering from the walls WW.

The resistance of these thick films was very low compared with that of the end-contacts, and hence was measured by comparing the potential drop across the film with that across a standard resistance in series with it. The current leads cc were connected to the graphited edges of the surface K as shown in fig. 1 *a*. The potential leads were attached by thin platinum wires dd to tiny glass beads fused to the surface K at the very edge of the graphiting. These beads were platinized and coated with graphite with the rest of the end-contacts as previously described (Lovell 1936 *b*).

In other respects the procedure of deposition and measurement was as described in the previous papers.

3. CALCULATION OF THE FILM RESISTIVITY

The width of the condensate was taken to be the mean width of the "umbra" and "penumbra" calculated from the geometry of the apparatus. With this particular slit system the overall width of the penumbra was 2·8 mm., of the umbra 2·4 mm., and it could readily be shown that the above approximation introduced a negligible error.

The length of the condensate was measured between the potential leads. In practice it proved impossible to place the potential leads on the geometrical edge of the graphited contacts, but measurements of the resistance of the graphiting indicated that no correction was needed on this account (at any rate for the thicker films). Where the film overlaps the graphiting the two are effectively in parallel, and the film resistance is very much lower than the resistance of the graphite undercoat.

The film thickness was calculated from the calibration of the beam as previously described.

4. RESULTS

Films of caesium, rubidium and potassium have been deposited on pyrex glass surfaces held at 64° K. The most significant results for these thick films are given in Table I. The values quoted for the bulk metal in this table are due to Hackspill (1910) (rubidium and caesium) and to Woltjer and Onnes (1924) (potassium).

Fig. 2 gives $\dfrac{\text{Film resistivity}}{\text{Bulk metal resistivity}}\left(\dfrac{\rho}{\rho_n}\right)$ against film thickness t for

TABLE I

	Bulk metal resistivity, ρ_n ohm cm.	Film resistivity, ρ ohm cm.	$(\rho_{90}-\rho_{64})$ bulk metal ohm cm.	$(\rho_{90}-\rho_{64})$ film ohm cm.
Caesium film thickness 10,280 A	$4{\cdot}00\times10^{-6}$	$4{\cdot}15\times10^{-6}$	$1{\cdot}36\times10^{-6}$	$1{\cdot}30\times10^{-6}$
Rubidium film thickness 7220 A	$2{\cdot}00\times10^{-6}$	$2{\cdot}80\times10^{-6}$	$0{\cdot}60\times10^{-6}$	$0{\cdot}68\times10^{-6}$
Potassium film thickness 2000 A	$1{\cdot}10\times10^{-6}$	$2{\cdot}85\times10^{-6}$	$0{\cdot}57\times10^{-6}$	$0{\cdot}59\times10^{-6}$

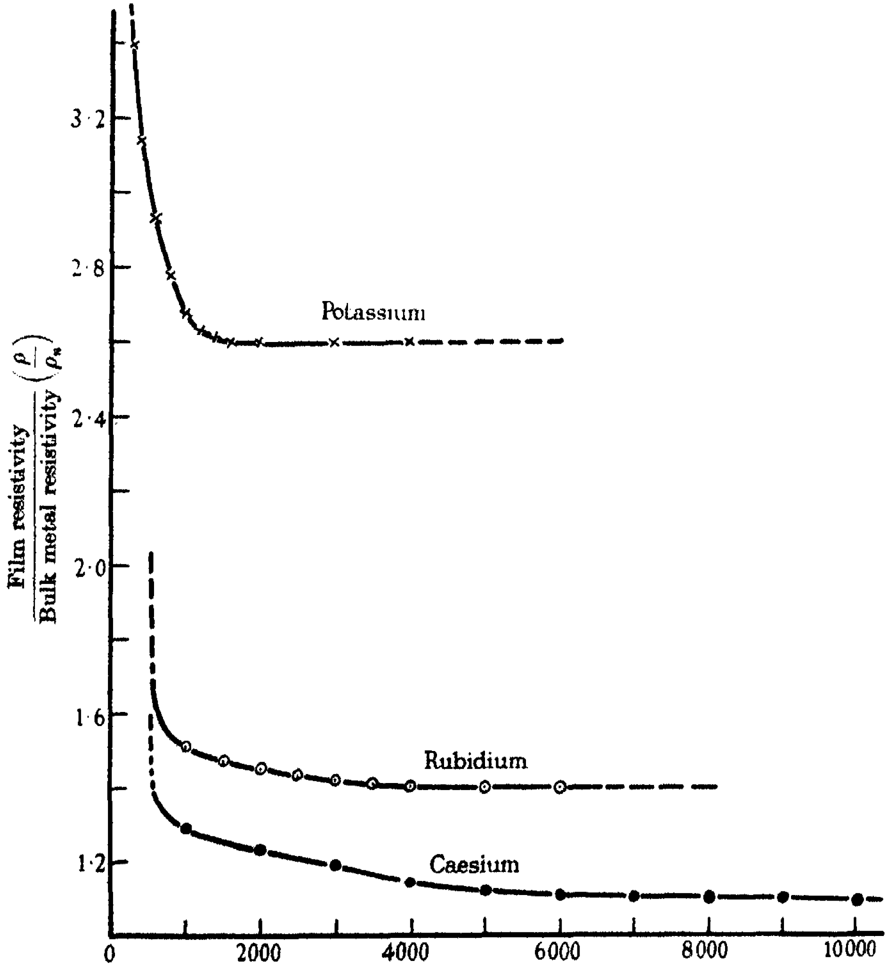


FIG. 2. Deposition of thick films of potassium (– × – × –), rubidium (– ⊙ – ⊙ –), and caesium (– ● – ● –) at 64° K.

$t > 500$ A. (The resistivities of the thinner films have been given in the previous papers.)

In the deposition of the caesium film the resistivity decreases steadily with thickness and converges to a more or less constant value at about 10,000 A, only 4 % in excess of that of the bulk metal. The change of resistivity in passing from 64 to 90° K. ($\rho_{90} - \rho_{64}$) is perfectly reversible and almost identical with that of the bulk metal. In the case of rubidium the resistivity does not decrease appreciably after about 4000 A and remains about 40 % greater than the value for the bulk metal. Potassium ceases to decrease in resistivity beyond 1600 A, and its resistivity is more than double that of the bulk metal.

5. Discussion

A comprehensive survey of the prominent factors influencing the resistance of metal films has recently been given by Appleyard (1937). There it is pointed out that films which agglomerate in their early stages of deposition must possess a finely polycrystalline structure, which may persist to great thicknesses. In such cases the condensate contains microcrystals possessing large surfaces of misfit, and the effect on the resistivity is to add a residual resistance to the normal value.

In previous work on the very thin films of the alkalis it has been shown that the initial layers increase in stability in the order: potassium, rubidium, caesium; and that at the lowest deposition temperature available (64° K.), the first layers of caesium are almost stable, while rubidium and potassium disintegrate considerably (Appleyard and Lovell 1937). On Appleyard's hypothesis we are thus led to expect quite a small residual resistance in the case of caesium, but successively greater for rubidium and potassium. Hence, we clearly have a highly probable explanation of the close agreement of the caesium film in Table I with the normal metal, and the greater disparity for rubidium and potassium.

One might attempt to explain the higher resistivity of the film material by assuming some form of geometrical disintegration of the film. The resistivity then obtained in these experiments would be merely a nominal value, say g times larger than the actual resistivity of the film material. g is a factor greater than unity and dependent only upon the degree of disintegration of the film. Thus the difference of resistivities between any two temperatures should also be g times the difference for the bulk metal, which is not consistent with experiment. Even if the cracks in the film are so close as to diminish the mean free path of the conduction electrons we should still expect the difference of the *absolute* resistivities to be greater than that for the bulk metal. We must therefore assume that these thick films differ from ordinary specimens of the metal only in being more finely polycrystalline.

There is one further observation strongly supporting this view. If the condensate really is an association of microcrystals, any annealing at a temperature higher than the temperature of deposition should result in a diminution of the film resistivity, due to relaxation in the surfaces of misfit (i.e. recrystallization). This has actually been observed consistently in these thick alkali films. For example, on the cessation of deposition at 64° K. a caesium film (10,200 A) had a resistivity of $4 \cdot 33 \times 10^{-6}$ ohm cm. After being raised to 90° K. the subsequent resistivity at 64° K. attained a steady value of $4 \cdot 15 \times 10^{-6}$ ohm cm. After this initial change no further alteration occurred and the change of resistivity with temperature up to 90° K. was perfectly reversible. Attempts to anneal these films at much higher temperatures have been frustrated by semi-disintegration of the film (see Appleyard 1937).

A similar hypothesis of relaxation of strain in the lattice has been invoked by Suhrmann and Barth (1936), and by Vand (1937), to explain resistivity changes on annealing. Aggregations of microcrystals have also been found by Fujiwara (1932) in thin molybdenum films; and transformations from the slightly strained to the normal lattice on annealing by Ogawa (1937).

Quantitatively, the resistivities of films less than 100 A in thickness showed good agreement with an expression derived from two assumptions: film thickness $t \ll \lambda_{\theta}$, the mean free path of the conduction electrons; and all collisions of the electrons with the boundaries of the film inelastic (Lovell 1936 *b*).

Over most of the present range of thicknesses the effect of the shortening of the mean free path by collisions with the boundaries is relatively unimportant. However, Fuchs (1938) has recently developed a more general expression for the resistivity of thin films with different proportions of elastic scattering at the boundaries. In his paper (to which we refer for details) the experimental results for caesium have been compared with the generalized expressions over the range from 0 to 10,000 A. The comparison suggests that the proportion of electrons scattered inelastically at the boundaries decreases from unity at very small thicknesses to about 0·3 or less at thicknesses greater than 150 A.

To summarize this work on the alkali metals, we may say that apparently all the phenomena can be explained in terms of a few familiar physical ideas without invoking any of the numerous special hypotheses which have been used by some previous workers in order to account for particular results (see Appleyard 1937 for bibliography). Finch (1937), reviewing experimental results on electron diffraction, likewise concludes that metal films on amorphous substrates do not differ fundamentally in structure from the normal metal. It seems possible therefore that a great deal of the inconsistency of previous work on other metals has been due to inadequate vacuum conditions and to insufficient attention to the surface condition of the substrate.

The majority of the work described here was carried out at intervals over an extended period in the H. H. Wills Physical Laboratory, University of Bristol, and the author is deeply indebted to Professor A. M. Tyndall for the necessary facilities, and to Dr Appleyard for invaluable advice. The author also desires to acknowledge the help given by Mr J. H. Burrow in the development of this high vacuum work, and his skill in the construction of the glass apparatus.

6. SUMMARY

The technique previously described for measuring the electrical conductivity of thin films deposited by evaporation, in very high vacua, on clean substrates, has been modified to include the deposition of thick films of the alkali metals.

Caesium films 10,000 A thick have resistivities only 4 % greater than the bulk metal, and identical temperature coefficients. Thick rubidium and potassium films have rather higher resistivities, but possess the same temperature coefficients as the bulk metal.

Evidence is produced that the films are polycrystalline, and that the higher resistivities are due to a simple residual resistance. The gradation in properties of the thick films is shown to follow consistently from the properties of the very thin films investigated previously.

From these results and other considerations it is concluded that there are no grounds for the belief that thin films differ essentially in structure from the normal metal.

REFERENCES

Appleyard, E. T. S. 1937 *Proc. Phys. Soc.* **49**, 118–35.
Appleyard, E. T. S. and Lovell, A. C. B. 1937 *Proc. Roy. Soc.* A, **158**, 718–28.
Finch, G. I. 1937 *Proc. Phys. Soc.* **49**, 113–17.
Fuchs, K. 1938 *Proc. Camb. Phil. Soc.* **34**, 100–8.
Fujiwara, T. 1932 *Mem. Coll. Sci. Engng Kyoto*, **15**, 31–3.

Hackspill, L. 1910 *C.R. Acad. Sci., Paris*, **151**, 305–8.
Lovell, A. C. B. 1936 *a* *Nature, Lond.*, **137**, 493.
— 1936 *b* *Proc. Roy. Soc.* A. **157**, 311–30.
Ogawa, S. 1937 *Sci. Rep. Tohoku Univ.* **26**, 92–105.
Suhrmann, R. and Barth, G. 1936 *Z. Phys.* **103**, 133–69.
Vand, V. 1937 *Z. Phys.* **104**, 48–67.
Woltjer, H. R. and Onnes, H. K. 1924 *4th Int. Congr. Refrig. Lond.* (*Commun. Phys. Lab. Leiden*, **173** *a*, 1–7).

The dispersion formula for nuclear reactions

By P. L. Kapur and R. Peierls

(*Communicated by R. H. Fowler, F.R.S.—Received* 17 *February* 1938)

1. Introduction

Bohr has shown that in a collision between two nuclei of which at least one is heavy, an unstable compound nucleus will be temporarily formed and that the lifetime of such a compound nucleus, measured on a nuclear scale, is usually very large. For this reason these compound nuclei have fairly well-defined energy levels ("virtual" or "resonance" levels), and the positions of these levels and the properties of the metastable states belonging to them determine the cross-sections for all nuclear reactions.

Breit and Wigner (1936) gave a formula for the probability of a nuclear reaction in terms of the virtual state, taking into account only one such state, while Bethe and Placzek (1937) and Bethe (1937) generalized the formula to take account of all possible virtual states. Their formula is usually referred to as the "dispersion formula" owing to its analogy with the formula for the dispersion of light by atomic systems.

In all these papers perturbation theory was used in order to derive the formula, and this involves the assumption that certain terms in the Hamiltonian of the nuclear system (such as the interaction between some of the constituents) may be considered as small, and that without these terms the virtual state would actually become stable. The perturbation theory was carried out only to the second order of approximation, and additional terms ought to arise when higher approximations are considered.

Now it is known that in actual fact all interaction forces in the nucleus are strong (this is essential to Bohr's theory), and therefore the dispersion formula might be expected to hold only in a very rough approximation. It therefore seemed worth while to investigate whether one cannot give a proof for this formula using only assumptions which are actually satisfied in the cases to which the formula is to be applied.

It is the purpose of this paper to show that this can be done, and we shall see that the dispersion formula can be justified provided only the widths of the resonance levels which contribute strongly to the process under consideration are smaller than the distances between the levels.

We start from the following consideration: If the levels are narrow, and therefore the escape of a particle from the nucleus is a rare event, the state of the compound nucleus will undergo very little change if we prevent the escape of the particle altogether, by imposing a suitable restriction on the wave function of the particle somewhere outside the nucleus. Thus, if we know the wave functions and energy levels of the system which is restricted by such a boundary condition, these will approximate to the solutions of the actual problem, and one can obtain the latter from the former by means of a kind of perturbation theory. This differs from the usual perturbation theory in that one treats a small change in the boundary conditions rather than a small change in the Hamiltonian of the system.

2. A SIMPLE EXAMPLE: THE ONE-BODY PROBLEM

Although the complete proof of the dispersion formula is simple in principle, it is obscured by the rather complicated notation which is necessary for the general case. We shall therefore illustrate the method by treating the case of the scattering of one particle in a central field of force, assuming this field to be wholly contained within a sphere of radius r_0. We further assume the wave-length of the particle to be so large that only the partial wave with $l = 0$ need be taken into account.

This partial wave $\phi = r\psi$ satisfies the wave equation

$$(E-H)\phi \equiv \frac{\hbar^2}{2m}\frac{d^2\phi}{dr^2}+[E-V(r)]\phi = 0, \tag{1}$$

in which $V(r)$ is the potential energy, and E the energy of the particle. For $r \geqq r_0$, V vanishes, and the equation reduces to

$$\frac{d^2\phi}{dr^2}+k^2\phi = 0, \tag{2}$$

where the wave vector k is defined by

$$k^2 = 2mE/\hbar^2. \tag{3}$$

The solution of (2) may be written

$$\phi = \frac{I}{k}\sin kr + Se^{ikr}, \tag{4}$$

where I is the amplitude of the incident wave and S that of the scattered wave. Evidently

$$\frac{d\phi}{dr} = I\cos kr + ikSe^{ikr}, \tag{5}$$

and hence we can express I and S by the value and slope of the function at r_0:

$$Ie^{-ikr_0} = \left(\frac{d\phi}{dr}\right)_{r_0} - ik\phi(r_0) \tag{6}$$

$$S = \cos kr_0 \,.\, \phi(r_0) - \frac{1}{k}\sin kr_0 \,.\left(\frac{d\phi}{dr}\right)_{r_0}. \tag{7}$$

From (6) it is seen that a state of affairs, in which *no incident wave* is present, would be described by the boundary condition

$$\frac{d\phi}{dr} - ik\phi = 0. \quad (r = r_0) \tag{8}$$

However, such a state of affairs is evidently impossible (unless ϕ vanishes throughout). The boundary condition (8) is not compatible with the wave equation (1). The condition (8), together with the requirement that ϕ vanishes at the origin in order to keep ψ finite, are two homogeneous boundary conditions at the ends of a finite interval, and it is known that with such boundary conditions the energy spectrum becomes discrete. Our energy value E is not an eigenvalue of this problem. We may, however, suppose we know the eigenvalues W_n of this problem. In other words, we suppose that we know the solutions ϕ_n of the equation

$$\frac{\hbar^2}{2m}\frac{d^2\phi_n}{dr^2} + [W_n - V(r)]\phi_n = 0 \tag{9}$$

with the boundary condition (8). It is important to realize that we consider k in (8) as a fixed constant, the same for all eigenvalues W_n. Thus for each value of k we obtain a different set of solutions ϕ_n and eigenvalues W_n.

The eigenvalues W_n are complex, owing to the fact that the boundary condition (8) contains i. Nevertheless, they form a complete set of orthogonal functions.

To show their orthogonality we multiply equation (9) by another function $\phi_{n'}$ and subtract the equation obtained by interchanging n and n'. After integration we obtain

$$(W_n - W_{n'})\int_0^{r_0} \phi_n \phi_{n'} dr + \frac{\hbar^2}{2m}\left[\phi_{n'}\frac{d\phi_n}{dr} - \phi_n \frac{d\phi_{n'}}{dr}\right] = 0.$$

The second term vanishes owing to the boundary condition (8), and it follows that

$$\int_0^{r_0} \phi_n \phi_{n'} dr = 0. \quad (W_n \neq W_{n'}) \tag{10}$$

It is important to note that the integrand in (10) is the direct product of the two functions, *not* the product of the one with the complex conjugate of the other.

Once the orthogonality has been established it follows in the usual way that the functions form a complete set, in the sense that any function ϕ which satisfies the boundary condition (8) and is sufficiently regular, can be expressed as a series

$$\phi = \Sigma a_n \phi_n. \tag{11}$$

Using these functions, we can then find an expression for the actual solution $\phi(r)$ which is a solution of the Schrödinger equation (1). It is true that this function does *not* satisfy the boundary condition (8) but instead the condition (6), and therefore cannot be expressed as a series of the type (11). But if we subtract from $\phi(r)$ any arbitrary function $\chi(r)$ which only satisfies the equation

$$\left(\frac{d\chi}{dr}\right)_{r_0} - ik\chi(r_0) = Ie^{-ikr_0}, \tag{12}$$

the remainder satisfies (8) and can therefore be expanded:

$$\phi = \sum_n a_n \phi_n + \chi. \tag{13}$$

We can now determine the coefficients a_n from the condition that the expansion (13) must satisfy the wave equation (1). Inserting in (1) and using (9), we easily obtain

$$(H_0 - E)\phi = \Sigma a_n (W_n - E)\phi_n + (H_0 - E)\chi = 0.$$

We multiply this equation by one particular ϕ_n and integrate. Owing to the orthogonality the sum then reduces to one term and we have

$$a_n(W_n - E)\int_0^{r_0} \phi_n^2\,dr + \int_0^{r_0} \phi_n(H_0 - E)\,\chi\,dr = 0. \tag{14}$$

The second term can be transformed into $\int \chi(H_0 - E)\phi\,dr$ by means of an integration by parts. This leaves two integrated terms, viz.

$$-\frac{\hbar^2}{2m}\left(\phi_n\frac{d\chi}{dr} - \chi\frac{d\phi_n}{dr}\right)_{r_0}.$$

Transforming these with the aid of (8) and (12), we obtain

$$(W_n - E)\,a_n\int_0^{r_0} \phi_n^2\,dr = \frac{\hbar^2}{2m}\phi_n(r_0)\,I e^{-ikr_0} - (W_n - E)\int_0^{r_0} \phi_n\chi\,dr. \tag{15}$$

The integral on the left-hand side might be made equal to unity by suitably normalizing the eigenfunction ϕ_n, but it seems more convenient to normalize the function in the usual way, by

$$\int_0^{r_0} |\,\phi_n\,|^2\,dr = 1, \tag{16}$$

and then

$$N_n = \int_0^{r_0} \phi_n^2\,dr \tag{17}$$

is a fixed constant of modulus less than unity, which is in general complex. For the integral on the right-hand side we introduce the abbreviation

$$b_n = \int_0^{r_0} \phi_n\chi\,dr, \tag{18}$$

but we shall see presently that it can be made as small as we like by suitably choosing χ. We have now

$$a_n = \frac{\hbar^2}{2m}\frac{\phi_n(r_0)\,I\,.\,e^{-ikr_0}}{(W_n - E)\,N_n} - \frac{b_n}{N_n}. \tag{19}$$

Thus, in principle, the function ϕ is known from (13) and we can therefore calculate the scattered amplitude from (7). By inserting (13) and using (8) and (12)

$$S = \Sigma a_n e^{-ikr_0}\phi_n(r_0) + e^{-ikr}\chi(r_0) - \frac{1}{k}I\sin kr_0 e^{-ikr_0}. \tag{20}$$

Inserting for a_n from (19)

$$S = I\left\{\sum_n \frac{\hbar^2}{2m}\frac{e^{-2ikr_0}\phi_n^2(r_0)}{(W_n - E)\,N_n} - \frac{1}{k}\sin kr_0 e^{-ikr_0}\right\} + e^{-ikr_0}\chi(r_0) - \sum_n \frac{b_n}{N_n}e^{-ikr_0}\phi_n(r_0). \tag{20a}$$

This formula must hold, whatever the function χ, provided only it satisfies (12). But we can make both the last terms in (20a) as small as we like; for example, by choosing

$$\chi(r) = \text{const.}\, e^{-\alpha(r_0-r)}.$$

We can satisfy (12) by choosing the value of the constant, and it is easy to see that if we let $\alpha \to \infty$ both the integral (18) and the value $\chi(r_0)$ will tend to zero.

Thus (20) becomes

$$S = \left\{\sum_n \frac{\hbar^2}{2m} \frac{e^{-2ikr_0}\phi_n^2(r_0)}{(W_n - E)\,N_n} - \frac{1}{k}\sin kr_0\, e^{-ikr_0}\right\} I. \tag{21}$$

Introducing further the abbreviation

$$u_n = \sqrt{(k\hbar^2/m)}\,\phi_n(r_0) \tag{22}$$

and writing

$$W_n = E_n - \tfrac{1}{2}i\gamma_n, \tag{23}$$

so that $\hbar/\gamma_n$ represents the "mean life" of the state n, the integral cross-section $\sigma = 4\pi\,|\,S/I\,|^2$ becomes

$$4\pi\left|\frac{S}{I}\right|^2 = \frac{\pi}{k^2}\left|\sum_n \frac{u_n^2\, e^{-2ikr_0}}{\{(E_n - E) - \frac{1}{2}i\gamma_n\}N_n} - 2\sin kr_0\, e^{-ikr_0}\right|^2. \tag{24}$$

The width γ_n of the level can be connected with the boundary value of the eigenfunction ϕ_n. If we multiply (9) by ϕ_n^* and, after subtracting the complex conjugate equation, integrate, we obtain

$$i\gamma_n\int_0^{r_0} |\,\phi_n\,|^2\,dr = \frac{\hbar^2}{2m}\left(\phi_n^*\frac{d\phi_n}{dr} - \phi_n\frac{d\phi_n^*}{dr}\right). \tag{24a}$$

Using the boundary condition (8), the normalization (16), and (22), this becomes

$$\gamma_n = |\,u_n\,|^2. \tag{25}$$

So far we have not made use of any approximating assumption. Our results, (24) and (25), are very similar to formulae (260) and (261) of Bethe (1937) as applied to pure scattering, if allowance is made for the term which Bethe and Placzek call "potential scattering" (Beth 1937, § 54 A). Our formula differs, however, in two respects: it contains the factor N_n in the denominator, and the quantity u_n is not in general real. Further, we do not know yet how u_n depends on k.

3. PERTURBATION THEORY FOR SMALL WIDTH OF THE LEVEL

Our results would be completely equivalent to the dispersion formula in its more usual form if we could show that (i) u_n is real, (ii) $N_n = 1$, and (iii) $|u_n|^2$ proportional to k.

One can see from (22) and (16), (17), that the conditions (i) and (ii) would be satisfied if $\phi_n(r)$ were real, while it follows from (22) that the last condition would be satisfied if $\phi_n(r_0)$ could be considered to be independent of k.

We shall now see that ϕ_n is indeed approximately real and independent of k, if the level is sufficiently narrow. First of all, it is quite evident that this will be the case if the width γ_n of the level actually vanishes. For then it follows from (25) that $\phi_n(r_0)$ must vanish, and from (8) that both $\phi_n(r_0)$ and $(d\phi_n/dr)_{r_0}$ are zero. In this limiting case it is, of course, irrelevant which value of k we use in (8) and we might replace (8) by the condition

$$\left(\frac{d\phi_n}{dr}\right)_{r_0} = 0. \tag{26}$$

This boundary condition is real and independent of k, and the same must then be true for the solution ϕ_n.

However, this argument does not tell us how narrow the level must actually be made in order for this approximation to apply. This may be decided in the following way: If the level is narrow we may expect that the solutions ϕ_n belonging to the boundary condition (8) will be similar to the functions Φ_n, say, which solve (9) with the boundary condition (26). These will be real functions and will belong to real eigenvalues E_n^0:

$$\frac{\hbar^2}{2m}\frac{d^2\Phi_n}{dr^2} + (E_n^0 - V)\Phi_n = 0, \tag{27}$$

$$\left(\frac{d\Phi_n}{dr}\right)_{r_0} = 0. \tag{27a}$$

Let now
$$\phi_n = \Phi_n + v_n, \tag{28}$$

and suppose v_n and
$$\epsilon_n = W_n - E_n^0 \tag{29}$$

to be small of first order. Then, neglecting terms of second order,

$$(H - E_n^0)\, v_n = \epsilon_n \Phi_n, \tag{30}$$

with
$$\left(\frac{dv_n}{dr}\right)_{r_0} = ik\Phi_n(r_0). \tag{31}$$

As before, we may expand
$$v_n = \Sigma a_m \Phi_m + \chi \tag{32}$$

with an arbitrary function χ which must satisfy

$$\left(\frac{d\chi}{dr}\right)_{r_0} = ik\Phi_n(r_0), \tag{33}$$

and which we can eventually eliminate. By applying to (30) the same transformation which led to (15) and leaving out at once the integral involving χ,

$$a_m(E_m^0 - E_n^0) = \epsilon_n \delta_{mn} + ik\frac{\hbar^2}{2m}\Phi_m(r_0)\Phi_n(r_0). \tag{34}$$

For $m = n$ this equation shows that

$$\epsilon_n \equiv E_n - E_n^0 - \tfrac{1}{2}i\gamma_n = ik\frac{\hbar^2}{2m}\Phi_n^2(r_0), \tag{35}$$

i.e. that there is no shift of the level in first approximation, and for $m \neq n$

$$a_m = ik\frac{\hbar^2}{2m}\frac{\Phi_m(r_0)\Phi_n(r_0)}{E_m^0 - E_n^0}. \tag{36}$$

In order to keep ϕ_n normalized, we put $a_m = 0$ for $m = n$. We have then from (36) and (32), dropping again a term in χ,

$$\phi_n(r_0) = \Phi_n(r_0) + ik\Phi_n(r_0)\frac{\hbar^2}{2m}\sum_m{}' \frac{\Phi_m^2(r_0)}{E_m^0 - E_n^0}. \tag{37}$$

Here the accent on the summation sign indicates that the term $m = n$ is to be omitted. Using (35), we see that the ratio between the imaginary and the real part is given by

$$\tfrac{1}{2}\sum_m{}' \frac{\gamma_m}{E_m^0 - E_n^0}. \tag{38}$$

In a similar way we can calculate N_n. Since $a_n = 0$ we evidently still have $N_n = 1$ in first approximation. If we write, up to second order,

$$\phi_n = \Phi_n + v_n + w_n,$$

we know that v_n is purely imaginary, while it is easy to prove that w_n is real. Hence if we further note that v_n is orthogonal to Φ_n, we have, up to terms of second order,

$$N_n = 1 + \int_0^{r_0} v_n^2\,dr + 2\int_0^{r_0} \Phi_n w_n\,dr, \tag{39}$$

whereas from the normalization condition

$$1 = 1 - \int_0^{r_0} v_n^2\,dr + 2\int_0^{r_0} \Phi_n w_n\,dr. \tag{40}$$

Using (40), the integral (39) becomes

$$N_n = 1 + 2\int_0^{r_0} v_n^2\,dr = 1 + 2\sum_m{}' a_m^2 = 1 - \tfrac{1}{2}\sum_m{}' \frac{\gamma_m \gamma_n}{(E_m - E_n)^2}.$$

It differs from unity by the term

$$1 - N_n = \tfrac{1}{2}\sum_m{}' \frac{\gamma_m \gamma_n}{(E_m - E_n)^2}. \tag{41}$$

The two sums, (38) and (41), are slightly different in character. (41) consists of positive terms only, and in order to make it small each separate term must be small compared to unity.

The sum (38), on the other hand, contains terms of both signs, and for this reason it might be small even when some of the individual terms are large. In that case the imaginary part of (37) would in first approximation be small. That, however, does not prove that the method of approximation converges. One can, in fact, show that the influence of the higher-order approximations will become large unless each separate term in (38) is smaller than unity, i.e. unless the width of the levels does not exceed their distance.†

Both sums, (38) and (41), converge at large m, owing to the fact that in the one-dimensional case which we are considering the density of eigenvalues is ultimately proportional to $E^{-\frac{1}{2}}$, but we shall have to take up this question again in the case of the many-body problem.

We may then summarize the results of this section by saying that the dispersion formula in the form given by Bethe and Placzek can be justified in our special example if the width of all levels is less than their distances apart.

Finally, we want to add a word about the choice of the radius r_0. The only assumption we have made about r_0 was that the potential should vanish for $r > r_0$, but this would be compatible with an arbitrarily large r_0. However, if we choose r_0 very large, the second term in (24), which represents the amplitude scattered by a hard sphere of radius r_0, would become large, and it would evidently have to be compensated by larger terms in the sum. Indeed, if r_0 is increased, the number of resonance levels is increased, and in particular there will now be more levels of very high energy. Thus the expression (24) is correct for any value of r_0, but it is more convenient for evaluation if r_0 is chosen as small as possible, i.e. so as just to enclose the potential field.

---

† The proof for this and a more detailed discussion of the case of a level width exceeding the distance apart will be given in a separate paper.

4. Necessary generalizations

The proof for the general case, which is applicable to actual nuclei, proceeds on very similar lines as that in the simplified example. Slight modifications are only necessary to allow for the coulomb field in the case of charged particles or the centrifugal force in the case of particles having a non-vanishing angular momentum, and also for the possible presence of magnetic forces inside the nucleus. In this section we shall outline the necessary modifications.

If the particle that collides with the nucleus is charged, or has an angular momentum, its wave equation is no longer given by (1), even outside the nucleus. This difficulty might be overcome by choosing the radius r_0 so large that beyond r_0 the coulomb field and the centrifugal force are negligible, but as explained at the end of the preceding section this would lead to a practically useless form of the dispersion formula, in which the result appears as a small difference between two large terms.

Instead, it is possible to proceed in the following way: At very large distances, the wave equation may certainly be replaced by (1),† and we may write its solutions in the form $\exp[\pm i(kr-\frac{1}{2}l\pi)]$, where l is the angular momentum. The second term in the exponent has been added merely for convenience.

Consider now the first of these two solutions which represents an outgoing spherical wave,

$$\phi_+ = e^{i(kr-\frac{1}{2}l\pi)}, \quad \text{(large } r\text{)} \tag{42a}$$

and introduce for its value and slope at r_0 the abbreviations

$$\phi_+(r_0) = A, \quad \left(\frac{d\phi_+}{dr}\right)_{r_0} = f.A. \tag{42b}$$

We have then for the second solution

$$\phi_-(r) = e^{-i(kr-\frac{1}{2}l\pi)}, \quad \text{(large } r\text{)} \tag{43a}$$

$$\phi_-(r_0) = A^*, \quad \left(\frac{d\phi_-}{dr}\right)_{r_0} = f^*A^*. \tag{43b}$$

It then follows immediately from the continuity equation as applied to either of the two functions that

$$(f-f^*)AA^* = 2ik. \tag{44}$$

If I_l is the contribution of the incident wave, and S_l that of the scattered spherical wave, the wave function is at large distances given by an expression which differs from (4) only by its phase

---

† Strictly speaking, this is not correct for a coulomb field, but it does apply to a screened coulomb field, which is sufficient for all practical purposes.

$$\phi = \frac{I_l}{k}\sin(kr - \tfrac{1}{2}l\pi) + S_l e^{i(kr-\frac{1}{2}l\pi)} \text{ at large } r. \tag{45}$$

Hence, by comparison with (42*a*, 43*a*),

$$\phi = \frac{I_l}{2ik}(\phi_+ - \phi_-) + S_l\phi_+, \tag{46}$$

and hence at r_0

$$\left.\begin{aligned} \phi(r_0) &= \frac{I_l}{2ik}(A - A^*) + S_l A, \\ \left(\frac{d\phi}{dr}\right)_{r_0} &= \frac{I_l}{2ik}(Af - A^*f^*) + S_l Af. \end{aligned}\right\} \tag{47}$$

Solving for I_l and S_l

$$\left(\frac{d\phi}{dr} - f\phi\right)_{r_0} = -\frac{I_l}{2ik}(f-f^*)A^*, \tag{48}$$

$$\left\{(A-A^*)\frac{d\phi}{dr} - (fA - f^*A^*)\phi\right\}_{r_0} = -S_l(f-f^*)AA^* = -2ikS_l$$

(the last equality from (44)).

The natural generalization of the boundary condition (8) is then

$$\left(\frac{d\phi}{dr} - f\phi\right)_{r_0} = 0, \tag{49}$$

and we shall now suppose we know the solutions of the Schrödinger equation with the boundary condition (49). We denote these solutions by ϕ_n as before, and by W_n the corresponding eigenvalues. Since f is complex, W_n will be complex and, denoting again by $-\frac{1}{2}\gamma_n$ its imaginary part, as in (23), we obtain from (24*a*) and (44)

$$i\gamma_n = \frac{\hbar^2}{2m}(f-f^*)(\phi_n^*\phi_n)_{r_0} = \frac{\hbar^2}{m}\frac{ik}{AA^*}(\phi_n^*\phi_n)_{r_0}. \tag{50}$$

The rest of the calculation proceeds as before. One sees from (48) that in the first term on the right-hand side of (15), I_l has to be replaced by $I_l(f-f^*)/2ik$. At the same time (48) shows that in the first term on the right-hand side of (20) we have to replace

$$e^{-ikr_0}\phi_n(r_0)$$

by

$$-\frac{1}{2ik}\left\{(A-A^*)\frac{d\phi_n}{dr} - (fA - f^*A^*)\phi_n\right\}_{r_0} = \frac{1}{2ik}(f-f^*)A^*\phi_n(r_0)$$

(cf. the boundary condition (49) for ϕ_n).

Then a repetition of the calculation which led to (20a) yields

$$\frac{S_l}{I_l} = -\frac{1}{2ik}\sum_n \frac{\hbar^2}{2m}\frac{\phi_n^2(f-f^*)}{(W_n-E)\,N_n}\frac{A^*}{A} + \frac{1}{2ik}\left(1-\frac{A^*}{A}\right), \tag{51}$$

and if we put

$$\sqrt{\left(\frac{\hbar^2}{2mi}(f-f^*)\right)}\,\phi_n(r_0) = u_n, \tag{52}$$

we have now

$$4\pi\left|\frac{S_l}{I_l}\right|^2 = \frac{\pi}{k^2}\left|-\Sigma\frac{u_n^2}{(E_n-E+\frac{1}{2}i\gamma n)\,N_n}\frac{A^*}{A}+\frac{1}{2i}\left(1-\frac{A^*}{A}\right)\right|^2. \tag{53}$$

Here the relation (25) holds again, because of (50) and (52), and therefore the first part of the formula is the same as in (24). The second term is different, but it represents again the "potential scattering", i.e. the amplitude that would be scattered by an impenetrable sphere of radius r_0, superimposed on the coulomb and centrifugal forces, because it is the scattered amplitude one would obtain from (14) by imposing the condition $\phi(r_0) = 0$.

Perturbation theory can be used as before to show that u_n is approximately real and independent of k, but we see from (50) that the width is now proportional to k/AA^* rather than to k, if ϕ_n may be considered as independent of k.

If the particle is uncharged, AA^* may easily be seen to be equal to $(kr_0)^{-2l}$, and then the width

$$\gamma_n = \frac{\hbar^2}{m}k^{2l+1}r_0^{2l}\,|\,\phi_n(r_0)\,|^2 \tag{54}$$

is proportional to k^{2l+1}. In the case of a charged particle, $1/AA^*$ contains also the transparency of the potential barrier.

The other modification we have to consider is due to the possibility of magnetic forces inside the nucleus, or other forces that will give rise to terms with imaginary coefficients in the differential equation; with these the derivation of (10) will no longer be applicable.

Instead, we consider, in addition to the functions ϕ_n which satisfy the boundary condition (49), new functions $\bar{\phi}_n$ which are the solutions of the Schrödinger equation with the conjugate complex boundary condition

$$\left(\frac{d\bar{\phi}_n}{dr} - f^*\bar{\phi}_n\right)_{r_0} = 0. \tag{55}$$

They belong to the eigenvalues W_n^*. Then instead of the orthogonality relation (10) we have

$$\int_0^{r_0}\bar{\phi}_n^*(r)\,\phi_{n'}(r)\,dr = 0, \quad \text{if } n \neq n'. \tag{56}$$

This result can easily be obtained† without reference to the detailed form of the Hamiltonian by applying the continuity equation to the function

$$\phi = \alpha\phi_n e^{-iW_n t/\hbar} + \beta\bar{\phi}_{n'} e^{-iW_{n'}^* t/\hbar} \tag{57}$$

which for any value of the coefficients α and β is a solution of the time-dependent wave equation. Therefore our previous results apply to this case if only we now define the constant N_n by

$$N_n = \int_0^{r_0} \phi_n^* \phi_n \, dr \tag{58}$$

instead of (17).

5. THE GENERAL CASE

For the general proof we shall assume that the energy of the incident particle is not sufficient to release a particle from the nucleus by an inelastic collision, or, in other words, that the energy of the virtual state which we are considering is insufficient to remove two particles from the nucleus. The results could without much difficulty be generalized to allow for the possibility of emission of two particles, but the formulae would then become too lengthy.

We assume further that the probability of more than one particle being found at a distance greater than r_0 from the centre of the nucleus is negligible. This is, with the above assumption about the energy, certainly correct for sufficiently large r_0, but it may imply choosing r_0 slightly larger than the actual radius of the nucleus.

If we denote (using as nearly as possible the notations of Bethe (1937)) the different particles that may possibly escape, by Q, the co-ordinates of the particle Q by x_Q, and its spin variable by s_Q, then for $r_Q > r_0$ we may expand the total wave function Ψ of the system into a series with respect to the stationary states q of the nucleus B which is left after the emission of Q:

$$\Psi = \sum_q \chi_q(B) \cdot \psi_q(x_Q s_Q). \quad r_Q \geqq r_0 \tag{59}$$

Then ψ_q must satisfy the differential equation

$$\nabla_Q^2 \psi_q(x_Q s_Q) + \left(k_{Qq}^2 - \frac{2M_{BQ}}{\hbar^2} \frac{Z_Q Z_B}{r_{QB}}\right) \psi_q(x_Q s_Q) = 0 \quad (r_Q \geqq r_0). \tag{60}$$

---

† The detailed proof will be carried out in the next section for the more general case.

Here M_{QB} is the reduced mass for the motion of the particle Q against the nucleus B, while k_{Qq} is defined by

$$k_{Qq}^2 = \frac{2M_{QB}}{\hbar^2}(E - E_Q - E_{Bq}), \tag{61}$$

where E is the total energy, E_{Bq} the energy of the nucleus B in the state q, and E_Q the internal energy (if any) of the particle Q. Analysing ψ_q in normalized spherical harmonics, we may write

$$\psi_q(x_Q s_Q) = \frac{1}{r_Q}\sum_{l,m}\phi_{qs_Qlm}(r_Q)\, Y_{lm}(\theta_Q\phi_Q). \tag{62}$$

We shall denote shortly as "the interior" that region of configuration space in which all particles have distances not exceeding r_0 from their common centre of gravity. This interior is then bounded by regions in which one of the distances is just equal to r_0 and where the boundary values of the wave function in the interior must fit on continuously to the expressions (59). For all particles except the incident one we have to impose the boundary condition that there be only outgoing waves at large distances, and according to the arguments of the preceding section this means imposing on ϕ_{qs_Qlm} the boundary condition

$$\left(\frac{d}{dr_Q} - f_{lqQ}\right)\phi_{qs_Qlm} = 0, \quad (r_Q = r_0) \tag{63}$$

where f is the quantity defined in (42) which in general depends on the angular momentum l and the wave vector, mass and charge of the particle Q and the residual nucleus B.

The boundary condition (63) does not apply if Q is the incident particle, but we shall, as before, suppose we know the solutions of the problem in which it holds on *all* boundaries. The eigenvalues of this problem we call again W_n, with

$$W_n = E_n - \tfrac{1}{2}i\gamma_n \tag{64}$$

and the eigenfunctions Ψ_n. At the boundary at which the distance of the particle Q is just equal to r_0, the expansion of the function Ψ_n,

$$\Psi_n = \sum_q \chi_q(B) \sum_{l,m} \frac{1}{r_Q}\phi^n_{qs_Qlm}(r_Q)\, Y_{lm}(\theta_Q\phi_Q),$$

must then, by definition, have the property that

$$\left(\frac{d}{dr_Q} - f_{lqQ}\right)\phi^n_{qs_Qlm} = 0. \quad (r_Q = r_0) \tag{65}$$

We define another set of functions $\tilde{\Psi}_n$ which satisfy the conjugate complex boundary condition

$$\left(\frac{d}{dr_Q}-f^*_{lqQ}\right)\tilde{\phi}^n_{qs_Qlm}=0, \quad (r_Q=r_0) \tag{66}$$

where $\tilde{\phi}$ is defined by an obvious generalization of (59) and (62). If $\tilde{W}_n$ denotes the corresponding eigenvalues,

$$\Psi_{\alpha\beta}=\alpha\Psi_n e^{-iW_nt/\hbar}+\beta\tilde{\Psi}_{n'}e^{-i\tilde{W}_{n'}t/\hbar}$$

is a solution of the time dependent Schrödinger equation. Hence it must satisfy the continuity equation

$$\frac{d}{dt}\int\limits_{\text{interior}}|\Psi_{\alpha\beta}|^2=\sum_Q\frac{\hbar^2}{2M_{QB}}\int\limits_{r_Q=r_0}\left(\Psi^*_{\alpha\beta}\frac{\partial\Psi_{\alpha\beta}}{\partial r_Q}-\Psi_{\alpha\beta}\frac{\partial\Psi^*_{\alpha\beta}}{\partial r_Q}\right)d\sigma_Q \tag{67}$$

The summation on the right-hand side extends over all particles that can possibly be emitted, and in each integral $d\sigma_Q$ denotes the integration with respect to all variables except r_Q. This relation must hold for all values of α and β, and if we equate the coefficients of $\alpha\beta^*$ we find that

$$(W_n-\tilde{W}^*_{n'})\int\limits_{\text{interior}}\Psi_n\tilde{\Psi}^*_{n'}\,d\tau=0.$$

Hence it follows that the integral must certainly vanish if the real parts of the eigenvalues are different:

$$\int\limits_{\text{interior}}\tilde{\Psi}^*_{n'}\Psi_n\,d\tau=0. \quad E_n\neq E_{n'} \tag{68}$$

If $n=n'$, we must either have

$$\tilde{W}^*_n=W_n, \tag{69}$$

or else the integral

$$N_n=\int\limits_{\text{interior}}\tilde{\Psi}^*_n\Psi_n\,d\tau \tag{70}$$

must vanish. It is clear that N_n will vanish only in exceptional cases and therefore we may assume (69) to hold.

Equating the coefficients of $\alpha\alpha^*$ in (67), we find, if we assume ϕ_n to be normalized by

$$\int\limits_{\text{interior}}|\Psi_n|^2\,d\tau=1,$$

the relation

$$\gamma_n=\sum_Q\sum_{l,m}\sum_{s,q}\frac{\hbar^2}{2M_{QB}}(f_{lqQ}-f^*_{lqQ})\,|\phi^n_{qslm}(r_0)|^2. \tag{71}$$

In analogy with Bethe (1937), equations (261), (261 *a*), we can write this in the form:

$$\gamma_n = \sum_{Qlmqs} \gamma^n_{Qlmqs}, \tag{72}$$

where
$$\gamma^n_{Qlmqs} = |u^n_{Qlmqs}|^2 = (f_{lqQ} - f^*_{lqQ}) \frac{\hbar^2}{2iM_{QB}} |\phi^n_{qslm}(r_0)|^2 \tag{73}$$

is the partial width of the level n which is due to the possible escape of the particle Q with angular momentum l, m and spin s, leaving the nucleus in the state q.

If we equate the coefficients of $\beta\beta^*$ in (67) we obtain in the same way

$$-i\tilde{\gamma}_n = \sum_Q \sum_{l,m} \sum_{s,q} \frac{\hbar^2}{2M_{QB}} (f_{lqQ} - f^*_{lqQ}) |\tilde{\phi}^n_{qslm}(r_0)|^2,$$

which, in connexion with (69), shows that

$$|\tilde{\phi}^n_{qslm}(r_0)|^2 = |\phi^n_{qslm}(r_0)|^2. \tag{74}$$

The two functions differ at the boundary only by a factor of modulus unity.

The actual wave function (59) describing the nuclear reaction in which the incident particle P collides with the nucleus A in the state p, satisfies the boundary condition (63) on all boundaries, except for $Q = P$ and $q = p$.

In the latter case we see from (48) that Ψ must satisfy, instead, the condition

$$\left(\frac{d}{dr_P} - f_{lpP}\right) \phi_{ps_plm} = -\frac{I_{lms}}{2ik_{Pp}} (f_{lpP} - f^*_{lpP}) A^*_{lpP}, \tag{75}$$

where I_{lms} is the partial amplitude of the incident wave belonging to the angular momentum l, m and spin s. If the incident wave is a plane wave in the z direction, carrying, in the centre of gravity system, one particle per unit area and per unit time, it is well known that

$$I_{lms} = \begin{cases} \left(\dfrac{M_{PA}(2l+1)}{4\pi\hbar k_{Pp}}\right)^{\frac{1}{2}} i^l, & m = 0 \\ 0, & m \neq 0 \end{cases} \tag{76}$$

We now write again
$$\Psi = \Sigma a_n \Psi_n + X, \tag{77}$$

where χ is an arbitrary function, subject only to the condition that it vanishes in the neighbourhood of all boundaries, except at that where $r_P = r_0$, and at the latter has the form

$$X = \frac{1}{r_P} \chi_p(A) \sum_{l,m} Y_{lm}(\theta_P \phi_P) \chi_{pslm}(r_P) \quad r_P \geqq r_0,$$

where A is the initial nucleus, χ_p its wave function in the initial state p. χ_{pslm} must satisfy the condition

$$\left(\frac{d}{dr_P}-f_{lpP}\right)\chi_{pslm} = -\frac{I_{lms}}{2ik_{Pp}}(f_{lpP}-f^*_{lpP})\,A^*_{lpP}, \tag{78}$$

according to (76). As before, all these conditions are compatible with making the function X in the interior and on all boundaries as small as we like, provided only its slope on the boundary $r_P = r_0$ keeps a finite value.

Then, from the condition that (77) is a solution of the wave equation we find in the same way as when deriving (19) that coefficient a_n has the value

$$a_n = \frac{1}{N_n}\frac{1}{(W_n-E)}\sum_{l,m,s}\frac{\hbar^2}{2M_{PA}}\frac{I_{lms}}{2ik_{Pp}}(f_{Ppl}-f^*_{Ppl})\,A^*_{Ppl}\tilde{\phi}^{*n}_{pslm}(r_0), \tag{79}$$

except for terms which contain X, but not its derivative. On the other hand, the partial wave representing the emission of the particle Q with angular momentum l', m' and spin s', leaving the nucleus B in the state q, is, according to (48),

$$S_{Qql'm's'} = -\frac{1}{2ik_{Qq}}\left\{\left[(A_{Qql'}-A^*_{Qql'})\frac{\partial}{\partial r_Q}-(f_{Qql'}A_{Qql'}-f^*_{Qql'}A^*_{Qql'})\right]\phi_{qs'l'm'}\right\}_{r_Q=r_0}. \tag{80}$$

For $Q, q \neq P, p$, this takes the form (cf. (77))

$$S_{Qql'm's'} = -\frac{1}{2ik_{Qq}}\sum_n a_n\{(f_{Qql'}-f^*_{Qql'})\,A^*_{Qql'}\phi^n_{qs'l'm'}(r_0)\},$$

and inserting the value of a_n from (79) and I_{lms} from (76)

$$S_{Qql'm's'} = -\frac{1}{2ik_{Qq}}(f_{Qql'}-f^*_{Qql'})\,A^*_{Qql'}\sum_n\frac{\phi^n_{qs'l'm'}}{N_n(W_n-E)}$$
$$\times\sum_{l,m,s}\frac{\hbar^2}{2M_{PA}}\frac{i^l}{2ik_{Pp}}(f_{Ppl}-f^*_{Ppl})\,.\,A^*_{Ppl}\tilde{\phi}^{*n}_{pslm}(r_0)\left(\frac{M_{PA}(2l+1)}{4\pi\hbar k_{Pp}}\right)^{\frac{1}{2}}.$$

This may be written as

$$S_{Qql'm's'} = \sum_{nls}\frac{1}{2}\left(\frac{(2l+1)\,M_{QB}}{4\pi\hbar k_{Qq}}\right)^{\frac{1}{2}}\frac{i^l}{k_{Pp}}\cdot\frac{u^n_{Qql'm's'}\tilde{u}^{*n}_{Pplos}}{N_n(E_n-E-\frac{1}{2}i\gamma_n)}, \tag{81}$$

if we define $$u^n_{Qqlms} = \left(\frac{\hbar^2}{4M_{QB}k_{Qq}}\right)^{\frac{1}{2}}\frac{f_{Qql}-f^*_{Qql}}{i}A^*_{Qql}\phi^n_{qslm}(r_0). \tag{82}$$

This definition of u agrees with the definition of $|u|^2$ contained in (73) because of (44).

Hence the integral cross section (number of emitted particles per unit time with unit incident intensity) becomes

$$\frac{\hbar k_{Qq}}{M_{QB}}\left| S_{Qql'm's'} \right|^2 = \frac{\pi}{k^2_{Pp}} \left| \sum_{n,l,s} \frac{i^l u^n_{Qql'm's'} \tilde{u}^{*n}_{Pplos} \cdot \sqrt{(2l+1)}}{(E_n - E - \frac{1}{2} i \gamma_n) N_n} \right|^2. \tag{83}$$

This equation has to be compared with the dispersion formula, for example, in the form given by Bethe (1937, equations (393) and (392)). One sees that the two formulae are identical (if allowance is made for the factor 2π in Bethe's definition of the matrix element (392)), except for the factor N_n and for the difference between ϕ_{pslm} and $\tilde{\phi}_{pslm}$.

For $Q = P$ and $q = p$, i.e. elastic scattering, one obtains the same formula, with the only difference that we have to add to the scattered amplitude a term which arises from the contribution of X to (80) and which is identical with the amplitude of the wave that would be scattered by an impenetrable sphere of radius r_0.

In order to obtain a cross section that can be immediately compared with the experiment, one ought to take into account the fact that degenerate states of the initial or the residual nucleus are not distinguished by the experiment. The generalizations which are necessary for this purpose are, however, completely identical with those discussed by Bethe and Placzek (1937), and therefore it seems unnecessary to repeat them in detail.

In order to show that the formula becomes completely identical with the usual dispersion formula if the levels are sufficiently narrow, we can again use perturbation theory as in § 3. The results of § 3 can easily be seen to be applicable to the general case, and it is again true that the phases of the expressions u and $\tilde{u}$ are nearly the same and that the factor N_n may be replaced by unity if the level width is less than their distance.

A further remark seems necessary with regard to the convergence of the sums of the type (38) which again represent the error committed in identifying u with $\tilde{u}$. In our present case the numerator in each term is given by the *partial* width of each level which is due to the emission of a particular particle with a particular angular momentum, leaving the residual nucleus in a particular state. Although the total level width increases with increasing energy of the level, this partial width will eventually decrease, because from a very high virtual state of the compound nucleus there is little chance of a particle escaping in such a way that the residual nucleus remains in a very low state of excitation. This is the reason why the sums converge in spite of the fact that the density of levels increases very rapidly with the excitation energy.

In conclusion, we would like to express our thanks to Dr L. Landau for stimulating conversations. The results of our § 2 are similar, although apparently not quite identical with results which Dr Landau obtained in a different way and mentioned to us before we started our calculations.

We understand from Professor Bohr that Dr Kalckar was considering a very similar method, but was prevented from completing the work by his sudden death.

SUMMARY

The dispersion formula of Bethe and Placzek, which expresses the cross-section for nuclear reactions in terms of the virtual levels of the compound nucleus and their widths, is derived without using the assumption that the motion of the incident particle within the nucleus can in any approximation be regarded as a one-body problem.

It is possible to derive a formula of the type of the dispersion formula without any assumption at all about the mechanism of the reaction. However, in the general case this formula contains more parameters than the usual dispersion formula and is therefore of less practical value.

The dispersion formula of Bethe and Placzek is shown to hold rigorously if only the widths of all virtual levels which contribute essentially to the expression are much smaller than the distances between the levels.

REFERENCES

Bethe, H. A. 1937 *Rev. Mod. Phys.* **9**, 69–244.
Bethe, H. A. and Placzek, G. 1937 *Phys. Rev.* **51**, 450.
Breit, G. and Wigner, E. 1936 *Phys. Rev.* **49**, 519.

A note on Guggenheim's theory of strictly regular binary liquid mixtures

By G. S. Rushbrooke, *St John's College, Cambridge*

(*Communicated by R. H. Fowler, F.R.S.—Received* 18 *February* 1938)

1. Introduction

This note attempts to revise and extend the formulae proposed by Guggenheim (1935) in his theory of a special (hypothetical) type of liquid mixture which, in the account of Guggenheim's paper given by Fowler (1936), is called strictly regular. According to the definitions in these sources a binary liquid mixture is said to be strictly regular if

(i) there is no volume change on mixing, whatever the relative amounts of the two components,

(ii) it possesses a definite co-ordination number, r say, so that each molecule, of either kind, in the liquid has precisely r (nearest) neighbours,

(iii) we may regard the potential energy of the mixture, for any arbitrary arrangement of the molecules consistent with (i) and (ii), as a sum of contributions from each pair of neighbouring molecules—the contribution being fixed when the molecules of the pair are named.

These conditions entail, *inter alia*, that the two species of molecule are each approximately spherical, equal or nearly equal in size, and have fields which are sensibly equal at distances greater than the distance between two neighbouring molecules.

We can then write the above potential energy (which can be called the configurational energy of the mixture) in the form

$$W = N_\alpha w_\alpha + N_\beta w_\beta + Xw, \tag{1}$$

where N_α and N_β are the numbers of the two sorts of molecule, α and β respectively, and rX is the number of $\alpha\beta$-pairs among all the pairs of neighbours in the liquid. This is equivalent to taking $2w_\alpha/r$, $2w_\beta/r$ and $2w_{\alpha\beta}/r$ as the contributions from each $\alpha\alpha$-, $\beta\beta$- and $\alpha\beta$-pair respectively, and then putting $2w_{\alpha\beta} - w_\alpha - w_\beta \equiv w$.

If $w = 0$ the solution is said to be perfect.

In effect (the notation is a little different, but only trivially so) Guggenheim, developing the statistical mechanics of solutions conforming to the above model, writes the partition function for the liquid mixture in the form

$$[J_\alpha(T)]^{N_\alpha}\,[J_\beta(T)]^{N_\beta}\,e^{-(N_\alpha w_\alpha+N_\beta w_\beta)/kT}\underset{\text{(different available configurations)}}{\int\dots\int} e^{-Xw/kT}\,(d\omega_\alpha)^{N_\alpha}\,(d\omega_\beta)^{N_\beta}, \qquad (2)$$

where $J_\alpha(T)$ is the partition function for the kinetic energy, internal vibrations and rotations of an α-molecule, etc., and the differentials in the integral refer to elements of volume.

If $w = 0$, i.e. if the strictly regular solution be perfect, the integral in (2) becomes

$$\underset{\text{(different available configurations)}}{\int\dots\int}(d\omega_\alpha)^{N_\alpha}\,(d\omega_\beta)^{N_\beta}, \qquad (3)$$

and is evaluated as $\quad v_\alpha^{N_\alpha}\,v_\beta^{N_\beta}\,(N_\alpha+N_\beta)!/(N_\alpha!\,N_\beta!),$

where v_α is a quasi-volume factor for an α-molecule—introduced as if each α-molecule were free to move in a small volume v_α without disturbing the other molecules of the liquid—v_β refers in the same way to a β-molecule, and the factor $(N_\alpha+N_\beta)!/N_\alpha!\,N_\beta!)$ takes account of the different rearrangements of the α- and β-molecules (or, more precisely, of their mean positions), the molecules of each kind being themselves indistinguishable.

But an appropriate form for the partition function is less evident if $w \neq 0$.

Guggenheim's purpose was to contrast the properties of a strictly regular solution, for which $w \neq 0$, with those of the perfect solution, with the same constitution, etc., but for which $w = 0$. As in his paper, the increments of the various thermodynamical functions, due to $w \neq 0$, will be denoted by ΔF, $\Delta E, \dots,$ where F, $E, \dots,$ stand for Helmholtz free energy, internal energy, etc.

If $\bar{\bar{X}}$ and $\bar{X}$ are defined by the formulae*

$$e^{-\bar{\bar{X}}w/kT}\int\dots\int(d\omega_\alpha)^{N_\alpha}\,(d\omega_\beta)^{N_\beta} = \int\dots\int e^{-Xw/kT}(d\omega_\alpha)^{N_\alpha}\,(d\omega_\beta)^{N_\beta},$$

$$\bar{X}\int\dots\int e^{-Xw/kT}(d\omega_\alpha)^{N_\alpha}\,(d\omega_\beta)^{N_\beta} = \int\dots\int Xe^{-Xw/kT}(d\omega_\alpha)^{N_\alpha}\,(d\omega_\beta)^{N_\beta},$$

and (like Guggenheim) we postulate that w is effectively independent of the temperature, then

$$\bar{X} = \bar{\bar{X}} - T\,\partial\bar{\bar{X}}/\partial T, \qquad (4)$$

---

* Guggenheim used the symbol $\bar{X}$ for our $\bar{\bar{X}}$: we prefer to keep to the usual convention that $\bar{X}$ denotes the average (equilibrium) value of X.

and, using hypothesis (i),

$$\left.\begin{aligned} \Delta G &= \Delta F = \bar{\bar{X}}w, \\ \Delta H &= \Delta E = \bar{X}w. \end{aligned}\right\} \tag{5}$$

But for these equations to be useful we must know how $\bar{\bar{X}}$ (or $\bar{X}$) depends on N_α, N_β and the temperature T. Guggenheim criticized the assumption (frequently made in thermodynamical work on the theory of solutions) that

$$\bar{\bar{X}}^2 = (N_\alpha - \bar{\bar{X}})(N_\beta - \bar{\bar{X}}), \quad (\bar{X} = \bar{\bar{X}})$$

because this implies perfectly random mixing of the molecules of the two components at all temperatures. He proposed instead the formula

$$\bar{\bar{X}}^2 = (N_\alpha - \bar{\bar{X}})(N_\beta - \bar{\bar{X}})\, e^{-2w/rkT},$$

guessing that this should be nearer the truth. Actually we shall show that

$$\bar{X}^2 = (N_\alpha - \bar{X})(N_\beta - \bar{X})\, e^{-2w/rkT} \tag{6}$$

is the correct equation, or rather, that this is the result we find on avoiding the difficulty of completely evaluating the partition function (2) by the indirect method first developed by Bethe (1935) in connexion with the theory of superlattices in metallic alloys. It is a first approximation only but should be sufficiently good for all practical purposes.

It seems preferable, for continuity, to develop the consequences of equation (6) a little at this point rather than after its validity has been established in § 2. On solving this equation we have

$$\bar{\bar{X}} = \frac{\{(N_\alpha + N_\beta)^2 - 4N_\alpha N_\beta (e^{2w/rkT} - 1)\}^{\frac{1}{2}} - (N_\alpha + N_\beta)}{2(e^{2w/rkT} - 1)},$$

to which a very good approximation, obtained by expanding the radical and retaining the first three terms, is

$$\bar{\bar{X}} = \frac{N_\alpha N_\beta}{N_\alpha + N_\beta}\left\{1 - \frac{N_\alpha N_\beta}{(N_\alpha + N_\beta)^2}(e^{2w/rkT} - 1)\right\}.$$

From this and equations (4) and (5) it follows that

$$\Delta H = \Delta E = \frac{N_\alpha N_\beta}{N_\alpha + N_\beta} w \left\{1 - \frac{N_\alpha N_\beta}{(N_\alpha + N_\beta)^2}(e^{2w/rkT} - 1)\right\},$$

$$\Delta G = \Delta F = \frac{N_\alpha N_\beta}{N_\alpha + N_\beta} w \left\{1 - \frac{N_\alpha N_\beta}{(N_\alpha + N_\beta)^2}\left(\frac{rkT}{2w} e^{2w/rkT} - \frac{rkT}{2w} - 1\right)\right\}.$$

These formulae give the required thermodynamical functions, and from them we could for instance investigate the existence of critical mixing conditions, i.e. values of the relative concentration N_α/N_β and the temperature T at which the liquid phase begins to separate into two parts, differing in constitution. In this case of course the above equations are to be interpreted as applying to each of these parts separately. However, we shall not do this here but shall derive the conditions for critical mixing directly from the statistical theory in § 3.

We shall continue, as in this introduction, to treat the liquid as incompressible so that its state is defined by N_α, N_β and T and does not depend on the external pressure. To extend the theory to include compressible fluids it is at least necessary to know how the total volume of the liquid depends on v_α and v_β. The physical significance of these quantities is not very certain. In this connexion we may perhaps conveniently digress a little here and discuss the above integration of the expression (3). If only one sort of molecule is present in the liquid phase Guggenheim's evaluation becomes simply $\int\dots\int(d\omega)^N = v^N$, where v is a quasi-volume factor for each molecule. Eyring and Hirschfelder (1937) (see also Lennard-Jones and Devonshire 1937) have criticized this on the grounds that it ascribes too crystalline a character to a liquid, and prefer to write $\int\dots\int(d\omega)^N = V_f^N/N!$, where V_f is a free volume which is equally available to all the molecules of the assembly: the $N!$ in the denominator is necessary because besides any one configuration this evaluation includes also others which differ only in the numbering of the molecules when $(d\omega)^N$ is written in full as $d\omega_1 d\omega_2 \dots d\omega_N$. Eyring and Hirschfelder evaluate V_f (in terms of the volume of the liquid and the size of the molecules) from a static model by writing $V_f = Nv_f$ and then taking v_f as the volume available to the centre of any molecule when this is treated as a rigid sphere imprisoned by its neighbours. If v_f is identified with v above then $V_f^N/N! = v^N e^N$, and Eyring and Hirschfelder's partition function is e^N times Guggenheim's. And of course there is the same possibility of difference in the case of a liquid mixture: if we suppose that V_f is available to both α- and β-molecules (treating them on an equal footing for simplicity), then the integral becomes $V_f^{N_\alpha+N_\beta}/(N_\alpha!\,N_\beta!)$, and this is

$$e^{N_\alpha+N_\beta}\,v_\alpha^{N_\alpha}\,v_\beta^{N_\beta}\,(N_\alpha+N_\beta)!/(N_\alpha!\,N_\beta!)$$

if we write $V_f = (N_\alpha+N_\beta)\,v_f$ and identify v_f with v_α and v_β. But the difference is significant only if we know the precise relation between the various quasi-volume factors and the total volume of the liquid phase—or associate them with others referring to a solid state—and for the present purpose such exponential factors can be regarded as modifying υ_α and υ_β or included in the factors $J_\alpha(T)$ and $J_\beta(T)$. It is of course necessary for the molecules to be able to change places, if we are to speak of an equilibrium state of the mixture (whence the factor $(N_\alpha+N_\beta)!/(N_\alpha!\,N_\beta!)$ in Guggenheim's evaluation of (3)), but in dealing with a strictly regular solution we must assume that these motions are not sufficiently frequent materially to upset the regular, co-ordinated, structure of the liquid.

Rather than introduce these quasi-volume factors it is perhaps more satisfactory, in the present instance, to suppose only that to a sufficient approximation we may construct the partition function for the liquid phase on the basis that each molecule of any one type makes the same contribution to it if we leave out of account the potential energy the molecule has, in virtue of its neighbours, when at the centre of its temporary cell—this potential being included as part of a configurational energy (1).

2. Statistical formulation and Bethe's method

We now proceed to the justification of formula (6) by an application of Bethe's indirect method which, for clarity, will be explained in detail. Since it is of interest to consider, with special reference to critical conditions, how the partial vapour pressures of such a strictly regular solution will depend on its constitution (relative concentration), we shall consider the liquid phase to be in equilibrium with a vapour phase (supposed perfect) and apply the analysis of statistical mechanics to the whole assembly. We suppose then that the states open to the molecules of the assembly are either those of a perfect gas or those of a strictly regular solution for which the partition function is (using the expressions (2), (3), etc.)

$$[J_\alpha(T)]^{N_\alpha}[J_\beta(T)]^{N_\beta}\sum_X t(rX, N_\alpha, N_\beta)\, e^{-(N_\alpha w_\alpha+N_\beta w_\beta+Xw)/kT}\, \upsilon_\alpha^{N_\alpha}\upsilon_\beta^{N_\beta},$$

or, with simpler notation,

$$\sum_X t(rX, N_\alpha, N_\beta)\, e^{-(N_\alpha w_\alpha+N_\beta w_\beta+Xw)/kT}\,[v_\alpha(T)]^{N_\alpha}\,[v_\beta(T)]^{N_\beta},$$

where $v_\alpha(T)$ and $v_\beta(T)$ are the partition functions, including the quasi-volume factors, for the motions of each α- and β-molecule respectively, and $t(rX, N_\alpha, N_\beta)$ is the number of distinguishable configurations of N_α α-molecules and N_β β-molecules, in the solution, for which there are rX $\alpha\beta$-pairs of neighbours. It is because an analytical expression for $t(rX, N_\alpha, N_\beta)$ is not known that we must proceed by an indirect method.

This partition function, $K(N_\alpha, N_\beta)$ say, refers to a liquid mixture containing N_α α-molecules and N_β β-molecules, which is only one example of the possible liquid phases: the one for which the vapour phase consists of M_α α-molecules and M_β β-molecules such that

$$N_\alpha + M_\alpha = X_\alpha; \quad N_\beta + M_\beta = X_\beta,$$

where X_α and X_β are the total numbers of α- and β-molecules, respectively, in the assembly. It is customary in such cases to determine the equilibrium example from the conditions

$$\left.\begin{aligned} \frac{\partial}{\partial N_\alpha} \log[K(N_\alpha, N_\beta)\,\lambda_\alpha^{N_\alpha}\lambda_\beta^{N_\beta}] = 0, \\ \frac{\partial}{\partial N_\beta} \log[K(N_\alpha, N_\beta)\,\lambda_\alpha^{N_\alpha}\lambda_\beta^{N_\beta}] = 0, \end{aligned}\right\} \tag{7}$$

where λ_α and λ_β are parameters which are then directly proportional to the partial vapour pressures of the α- and β-molecules. (The conditions (7) ensure that the number of complexions pertaining to any particular example is a maximum for the example specified by N_α and N_β.) But in the present instance this procedure is not feasible because we cannot form $K(N_\alpha, N_\beta)$ without knowing $t(rX, N_\alpha, N_\beta)$: it is necessary to replace the above equations by others, equivalent to them but in a form suitable for using Bethe's method. We therefore consider the partition function, $K(N_\alpha + N_\beta)$ say, for an example of the assembly which is now specified only in that the total extent of the liquid phase $N_\alpha + N_\beta$, is prescribed—but not its constitution (relative concentration) N_α/N_β. This is given by

$$K(N_\alpha + N_\beta) = \sum_{N_\alpha/N_\beta} \sum_X t(rX, N_\alpha, N_\beta)\, e^{-(N_\alpha w_\alpha + N_\beta w_\beta + Xw)/kT} [v_\alpha(T)]^{N_\alpha} [v_\beta(T)]^{N_\beta},$$

where $\sum\limits_{N_\alpha/N_\beta}$ denotes summation over all possible values of N_α and N_β for which $N_\alpha + N_\beta$ is constant. The equilibrium example of the assembly will now be determined by the condition

$$\frac{\partial}{\partial (N_\alpha + N_\beta)} \log Q = 0,$$

and its properties given by the formulae

$$\left.\begin{aligned} \bar{N}_\alpha = \lambda_\alpha \frac{\partial}{\partial \lambda_\alpha} \log Q, \\ \bar{N}_\beta = \lambda_\beta \frac{\partial}{\partial \lambda_\beta} \log Q, \end{aligned}\right\} \tag{8}$$

and (symbolically)
$$\bar{X} = -kT\frac{\partial}{\partial w}\log Q, \tag{9}$$

where Q denotes the expression obtained when $\lambda_\alpha v_\alpha(T)$ and $\lambda_\beta v_\beta(T)$ replace $v_\alpha(T)$ and $v_\beta(T)$, respectively, in $K(N_\alpha + N_\beta)$.

Of course these equations must determine the same equilibrium state for the assembly as do equations (7), in fact it is easy to verify this for the simple case when $w = 0$, but their importance lies in their showing that N_α, N_β and X have, for this equilibrium state, the values we should obtain by averaging over all different arrangements of molecules in the liquid phase such that this has its equilibrium extent, $\bar{N}_\alpha + \bar{N}_\beta$, assigning a statistical weight

$$e^{-(N_\alpha w_\alpha + N_\beta w_\beta + Xw)/kT}\,(\lambda_\alpha v_\alpha(T))^{N_\alpha}\,(\lambda_\beta v_\beta(T))^{N_\beta} \tag{10}$$

to every arrangement in which there are N_α α-molecules, N_β β-molecules and rX $\alpha\beta$-pairs of neighbours. We cannot sum (10) over all arrangements with given $N_\alpha + N_\beta$ to construct Q, because we do not know $t(rX, N_\alpha, N_\beta)$, but this difficulty can now be avoided by using Bethe's method in which we consider explicitly only the ways of covering a central site in the liquid and its first shell, i.e. the r neighbouring places nearest to it. For the results to be physically significant we have also in some way to take account of the other liquid molecules (outside these $r+1$): physical ideas will afford a check on the approximations.

It may perhaps seem that if we are going to consider explicitly only a small patch of the liquid we ought to replace the exponential term in (10) by the expression from which it was initially derived, and rewrite the statistical weight in the form

$$x_{\alpha\alpha}^{\frac{1}{2}r(N_\alpha - X)}\,x_{\beta\beta}^{\frac{1}{2}r(N_\beta - X)}\,x_{\alpha\beta}^{rX}\,(\lambda_\alpha v_\alpha(T))^{N_\alpha}\,(\lambda_\beta v_\beta(T))^{N_\beta},$$

where $x_{\alpha\alpha} = e^{-2w_\alpha/rkT}$ = factor for each $\alpha\alpha$-pair of neighbours,

$x_{\beta\beta} = e^{-2w_\beta/rkT}$ = factor for each $\beta\beta$-pair of neighbours,

and $x_{\alpha\beta} = e^{-2w_{\alpha\beta}/rkT}$ = factor for each $\alpha\beta$-pair of neighbours.

Actually we shall see that this is unnecessary, (10) could be used as it stands just as well, but to start with we will introduce the factors separately and so, when forming the statistical weight for any particular arrangement of molecules on the $r+1$ sites, include factors $\lambda_\alpha v_\alpha(T)$ for each α-molecule, $\lambda_\beta v_\beta(T)$ for each β-molecule, and $x_{\alpha\alpha}, x_{\beta\beta}$ and $x_{\alpha\beta}$ for each $\alpha\alpha$- $\beta\beta$- and $\alpha\beta$-pair of neighbours, respectively, in this arrangement. A statistical weight formed of these alone would entirely ignore all the rest of the liquid phase, surrounding the small patch of $r+1$ molecules. We have yet to say how to take account of the influence of this environment.

Before approximating at all it is evident that

$$t(rX, N_\alpha, N_\beta)\, e^{-(N_\alpha w_\alpha + N_\beta w_\beta + Xw)/kT}\, (\lambda_\alpha v_\alpha(T))^{N_\alpha}\, (\lambda_\beta v_\beta(T))^{N_\beta}$$

is the statistical weight of the class of all configurations for which there are N_α α-molecules, N_β β-molecules and rX $\alpha\beta$-pairs of neighbours.

Case 1. $w > 0$

If $w > 0$ then, mathematically, this means that at low temperatures the classes of configurations with the greatest statistical weights, for given N_α and N_β, will be those for which X is small: for when T is small the factor $e^{-Xw/kT}$ has more influence than the factor $t(rX, N_\alpha, N_\beta)$ on the magnitude of the statistical weight.

Physically $w > 0$ means that the smaller the number of $\alpha\beta$-pairs the less is the potential energy of the mixture and we therefore expect that at low temperatures the liquid phase will consist of a dilute solution of α in β and a dilute solution of β in α.

Clearly the mathematical and physical interpretations are equivalent. The environment therefore is one which, speaking loosely, gives preference to either α- or β-molecules in the first shell; we therefore introduce into the statistical weight, as already formed for any particular arrangement of molecules on the $r + 1$ sites, a factor ϵ for each β-molecule (say) in the first shell.

We do not introduce two ϵ's, ϵ_α for each α-molecule and ϵ_β for each β-molecule (in the first shell), because since the total number of molecules in the first shell, α's and β's together, is always r, only the ratio of these ϵ's would matter when we used the statistical weights for calculating averages (equilibrium properties)—that this is so can be verified immediately by trial. The ϵ that we introduce acts as a preference factor for α- or β-molecules according as it is < 1 or > 1.

And of course we do not introduce an ϵ factor for the central molecule when this is a β-molecule, because we are taking account of its interactions, with the first shell molecules, explicitly, when we include the various x factors: it is only because we do not know how otherwise to take account of the influence on the first shell molecules of the rest of the liquid that we introduce the ϵ factors. The supposition that the effect of this environment is to give preference to α- or β-molecules in the first shell is in accord with the above interpretation (mathematical or physical) of the hypothesis $w > 0$.

Case II. $w < 0$

Mathematically this means that, at low temperatures, the classes of configurations with the greatest statistical weights will be those for which X is great: for, as in case I, when T is small the factor $e^{-Xw/kT}$ has more influence than the factor $t(rX, N_\alpha, N_\beta)$ on the magnitude of the statistical weight.

Physically $w < 0$ implies that the greater the number of $\alpha\beta$-pairs the less will be the potential energy of the mixture, and so at low temperatures there will be nearly the maximum number of $\alpha\beta$-pairs.

Bethe has discussed the case $N_\alpha = N_\beta$ for alloys of the β-brass type and shown that for $T < T_c$, where T_c is a critical temperature, there is superlattice structure and long-distance order: in the perfect superlattice every α-molecule is completely surrounded by β-molecules, and vice versa. The effect of the external molecules is then to make those in the first shell more likely to be "right" than "wrong", and ϵ's have to be introduced according as the first shell molecules are "right" or "wrong" irrespective of whether they are α's or β's. Owing to the particularly simple nature of the superlattice structure in this case, there is no difficulty in taking explicit account of the interactions between the central molecule and those in its first shell. Easthope (1937) has extended this theory to cases when N_α and N_β are not equal, provided that other types of superlattice are not formed. But it is not evident how to proceed in the general case when $N_\alpha \neq N_\beta$.

We must beware, however, of pushing our semi-crystalline model for the liquid phase too far. Bethe's theory referred to metallic alloys. Superlattice structure in a liquid mixture is most improbable, for, on account of the nature of the liquid state, we must expect that slow movements of the lattice points (about which the molecules vibrate) or points of abnormal co-ordination will destroy the regularity necessary for there to be long-distance order. If we discount the possibility of long-distance order we have no need to introduce the conception of "right" and "wrong" molecules and can suitably take account of the external interactions by introducing the ϵ's for α's or β's in the first shell, as in case I.

We proceed on this basis to construct the statistical weights and calculate average (equilibrium) properties of the liquid and vapour phases.

We have first to specify the geometry of the $r + 1$ points which represent the patch of sites in the liquid phase, and shall suppose that this is such that none of the neighbours of the central site (i.e. the first shell sites) are neighbours of each other. The implications of this simplification, which means that the only interactions considered explicitly are those between the central molecule and its neighbours, will be discussed briefly at the end of this section.

Then we find:

The statistical weight of the set of all configurations in which the central site is occupied by an α-molecule is

$$\lambda_\alpha v_\alpha(T)\left\{\sum_{n=0}^{r}\binom{r}{n}x_{\alpha\alpha}^{r-n}x_{\alpha\beta}^{n}\epsilon^{n}(\lambda_\alpha v_\alpha(T))^{r-n}(\lambda_\beta v_\beta(T))^{n}\right\}$$
$$=\lambda_\alpha v_\alpha(T)\{x_{\alpha\alpha}\lambda_\alpha v_\alpha(T)+\epsilon x_{\alpha\beta}\lambda_\beta v_\beta(T)\}^r\equiv Q^*_\alpha,$$

and similarly

The statistical weight of the set of all configurations in which the central site is occupied by a β-molecule is

$$\lambda_\beta v_\beta(T)\left\{\sum_{n=0}^{r}\binom{r}{n}x_{\beta\beta}^{r-n}x_{\alpha\beta}^{n}\epsilon^{r-n}(\lambda_\alpha v_\alpha(T))^{n}(\lambda_\beta v_\beta(T))^{r-n}\right\}$$
$$=\lambda_\beta v_\beta(T)\{x_{\alpha\beta}\lambda_\alpha v_\alpha(T)+\epsilon x_{\beta\beta}\lambda_\beta v_\beta(T)\}^r\equiv Q^*_\beta,$$

so that the approximate expression by which we intend to replace Q is

$$Q^* = Q^*_\alpha + Q^*_\beta,$$

but now equations (8) and (9) are replaced by

$$\left.\begin{aligned}\overline{N}_\alpha &= \frac{(\overline{N}_\alpha+\overline{N}_\beta)}{r+1}\lambda_\alpha\frac{\partial}{\partial\lambda_\alpha}\log Q^*,\\ \overline{N}_\beta &= \frac{(\overline{N}_\alpha+\overline{N}_\beta)}{r+1}\lambda_\beta\frac{\partial}{\partial\lambda_\beta}\log Q^*,\end{aligned}\right\}\tag{11}$$

and[†]

$$\overline{X} = \tfrac{1}{2}(\overline{N}_\alpha+\overline{N}_\beta)\frac{1}{r}x_{\alpha\beta}\frac{\partial}{\partial x_{\alpha\beta}}\log Q^*,\tag{12}$$

since Q^* refers explicitly to only $r+1$ sites. Of course these equations are essentially formal, and in them ϵ must be treated as a constant.

We have yet to say how to fix the value to be given to ϵ. We find this from the condition that there shall be no real difference between the central site and any of the first shell sites. Thus the statistical weight of the set of all configurations in which some particular first shell site is occupied by an α-molecule must be the same as the statistical weight of the set of all configurations in which the central site is occupied by an α-molecule.

[†] $x_{\alpha\beta}\,\partial/\partial x_{\alpha\beta}\log Q^*$ gives the average number of $\alpha\beta$-bonds including one particular site—the central site—and to get the total number of $\alpha\beta$-pairs ($=rX$) we have to multiply this by the total number of sites and divide by two since otherwise each $\alpha\beta$-pair would be counted twice.

Whether we express the equivalence of the sites in terms of α- or β-molecules we shall get the same† ϵ. The equivalence for α-molecules gives the equation

$$\begin{aligned}(\lambda_\alpha v_\alpha(T))^2 x_{\alpha\alpha}(x_{\alpha\alpha}\lambda_\alpha v_\alpha(T)+\epsilon x_{\alpha\beta}\lambda_\beta v_\beta(T))^{r-1}&\\ +\lambda_\alpha v_\alpha(T)\lambda_\beta v_\beta(T)\,x_{\alpha\beta}(x_{\alpha\beta}\lambda_\alpha v_\alpha(T)+\epsilon x_{\beta\beta}\lambda_\beta v_\beta(T))^{r-1}&\\ =\lambda_\alpha v_\alpha(T)\,(x_{\alpha\alpha}\lambda_\alpha v_\alpha(T)+\epsilon x_{\alpha\beta}\lambda_\beta v_\beta(T))^{r},&\end{aligned}$$

i.e.

$$\epsilon=\left(\frac{x_{\alpha\beta}\lambda_\alpha v_\alpha(T)+\epsilon x_{\beta\beta}\lambda_\beta v_\beta(T)}{x_{\alpha\alpha}\lambda_\alpha v_\alpha(T)+\epsilon x_{\alpha\beta}\lambda_\beta v_\beta(T)}\right)^{r-1}. \tag{13}$$

Using this equation for ϵ after performing the formal operations in (11) yields the formulae

$$\bar{N}_\alpha=(\bar{N}_\alpha+\bar{N}_\beta)\,Q^*_\alpha/Q^*,$$
$$\bar{N}_\beta=(\bar{N}_\alpha+\bar{N}_\beta)\,Q^*_\beta/Q^*,$$

and equation (12) may be written

$$\bar{X}=\tfrac{1}{2}x_{\alpha\beta}\left(\bar{N}_\alpha\frac{\epsilon\lambda_\beta v_\beta(T)}{x_{\alpha\alpha}\lambda_\alpha v_\alpha(T)+\epsilon x_{\alpha\beta}\lambda_\beta v_\beta(T)}+\bar{N}_\beta\frac{\lambda_\alpha v_\alpha(T)}{x_{\alpha\beta}\lambda_\alpha v_\alpha(T)+\epsilon x_{\beta\beta}\lambda_\beta v_\beta(T)}\right).$$

Combining these equations we get

$$\frac{\bar{N}_\alpha}{\bar{N}_\beta}=\frac{\lambda_\alpha v_\alpha(T)\,\{x_{\alpha\alpha}\lambda_\alpha v_\alpha(T)+\epsilon x_{\alpha\beta}\lambda_\beta v_\beta(T)\}^r}{\lambda_\beta v_\beta(T)\,\{x_{\alpha\beta}\lambda_\alpha v_\alpha(T)+\epsilon x_{\beta\beta}\lambda_\beta v_\beta(T)\}^r}, \tag{14}$$

and

$$\left.\begin{aligned}\bar{X}&=\bar{N}_\alpha\,\epsilon x_{\alpha\beta}\lambda_\beta v_\beta(T)/\{x_{\alpha\alpha}\lambda_\alpha v_\alpha(T)+\epsilon x_{\alpha\beta}\lambda_\beta v_\beta(T)\}\\ &=\bar{N}_\beta\,x_{\alpha\beta}\lambda_\alpha v_\alpha(T)/\{x_{\alpha\beta}\lambda_\alpha v_\alpha(T)+\epsilon x_{\beta\beta}\lambda_\beta v_\beta(T)\},\end{aligned}\right\} \tag{15}$$

formulae which are almost self-evident and which include all that is expressed in the equations (11), (12) and (13).

From equations (15) it follows at once that

$$(\bar{N}_\alpha-\bar{X})\,(\bar{N}_\beta-\bar{X})=\bar{X}^2x_{\alpha\alpha}x_{\beta\beta}/x^2_{\alpha\beta},$$

i.e.

$$\bar{X}^2=(\bar{N}_\alpha-\bar{X})\,(\bar{N}_\beta-\bar{X})\,e^{-2w/rkT},$$

which is equation (6). It is also apparent from (14) and (15) that we could equally well have used the expression (10) in its original form: a transformation to the variables

---

† It is easy to show that the equivalence condition must be simply the equation $\lambda_\alpha\frac{\partial}{\partial\lambda_\alpha}Q^*_\beta=\lambda_\beta\frac{\partial}{\partial\lambda_\beta}Q^*_\alpha$. This is no way depends on the specification that no two of the neighbours of any one site are neighbours of each other.

$$\xi_\alpha = \lambda_\alpha v_\alpha(T)\, x_{\alpha\alpha}^{r/2}; \quad \xi_\beta = \lambda_\beta v_\beta(T)\, x_{\beta\beta}^{r/2};$$

$$\epsilon' = \epsilon x_{\beta\beta}^{-(r-1)/2}\, x_{\alpha\alpha}^{(r-1)/2} \quad \text{and} \quad x = x_{\beta\beta}^{-\frac{1}{2}}\, x_{\alpha\alpha}^{-\frac{1}{2}}\, x_{\alpha\beta} = e^{-w/rkT},$$

puts (14) in the form

$$\frac{\overline{N}_\alpha}{\overline{N}_\beta} = \frac{\xi_\alpha(\xi_\alpha + \epsilon' x \xi_\beta)^r}{\xi_\beta(x\xi_\alpha + \epsilon' \xi_\beta)^r}, \tag{14'}$$

and the other equations also take the forms we should then have obtained, for instance

$$\epsilon' = \left(\frac{\xi_\alpha x + \epsilon' \xi_\beta}{\xi_\alpha + \epsilon' x \xi_\beta}\right)^{r-1}. \tag{13'}$$

In this discussion (cf. (11) and (12)) we have, effectively, considered one homogeneous phase; but we shall see that sometimes the equations have two solutions and indicate that the liquid mixture splits into two phases in mutual equilibrium. In that case equation (6) applies separately to each of these phases: and it is easy to show that the configurational energy per unit volume is the same for each of them.

The above equations are valid, as a first approximation, provided that the molecules of the liquid mixture are not too close packed: for in that case first shell sites are neighbours of each other and for as good an approximation we need to include, explicitly, the interactions between neighbouring first shell molecules. This leads to more complicated equations from which equation (6) does not follow. A short account of some calculations for close packing (face-centred cubic) will be given in § 4; an elaborate discussion would involve considerable numerical work and is, perhaps, hardly appropriate since in any case we do not know how justified we are in treating a liquid phase as an assembly of molecules with a precise co-ordination number r. But we shall next consider what, on the simple theory above, will be the nature of the partial vapour pressure curves.

3. PARTIAL VAPOUR PRESSURES, AND CRITICAL-SOLUTION-TEMPERATURE

For the vapour phase (supposed perfect) we have the usual equations

$$M_\alpha = \lambda_\alpha (2\pi m_\alpha kT)^{3/2}\, V b_\alpha^{(g)}(T)/h^3,$$

$$M_\beta = \lambda_\beta (2\pi m_\beta kT)^{3/2}\, V b_\beta^{(g)}(T)/h^3,$$

where m_α and m_β are the masses of α- and β-molecules respectively, the $b^{(g)}(T)$ terms are partition functions for their internal motions (if any) in the vapour phase and V is the volume of this phase. They show that λ_α and λ_β, and so also ξ_α and ξ_β above, are proportional, at any given temperature, to the partial vapour pressures of α- and β-molecules respectively.

Equations (13′) and (14′) do not enable us to determine these ξ_α, ξ_β directly but only their ratio. We shall now consider how this ratio behaves as a function of the relative concentration N_α/N_β of the liquid phase, at different constant temperatures T.

Writing $N_\alpha/(N_\alpha+N_\beta) = \theta_\alpha$; $N_\beta/(N_\alpha+N_\beta) = \theta_\beta$ and putting $\xi_\beta/\xi_\alpha = \xi$ and $\xi\epsilon' = \eta$, equations (14′) and (13′) become

$$\theta_\alpha = \frac{1+x\eta}{1+2x\eta+\eta^2}; \quad \theta_\beta = \frac{(x+\eta)\,\eta}{1+2x\eta+\eta^2},$$

and
$$\xi = \eta\left(\frac{1+x\eta}{x+\eta}\right)^{r-1}. \tag{13''}$$

$\partial\xi/\partial\theta_\alpha$ vanishes if $\partial\xi/\partial\eta$ vanishes ($\partial\theta_\alpha/\partial\eta \neq 0$ because $x>0$ and $\eta>0$), and $\partial\xi/\partial\eta = 0$ implies $1/(r-2)\,\eta = (1-x^2)/x(1+2x\eta+\eta^2)$. On eliminating η in favour of θ_α this becomes

$$\left(\theta_\alpha + \frac{1}{r-2}\right)\left(1-\theta_\alpha + \frac{1}{r-2}\right) = \frac{1}{x^2}\theta_\alpha(1-\theta_\alpha),$$

an equation which has real roots in θ_α only if $x \leqslant 1-2/r$. If this condition is satisfied then $\partial\xi/\partial\theta_\alpha = 0$ for two real values of θ_α which are at equal distances from $\theta_\alpha = \frac{1}{2}$. If the condition is not satisfied then ξ is a monotonic function of θ_α.

Although we have so far dealt only with ξ, the results also describe the behaviour of ξ_α and ξ_β, and thus the partial vapour pressure curves (which show how the partial vapour pressures change as the relative concentration N_α/N_β of the solution is altered, at different constant temperatures). For it is well known that these curves are connected by the Gibbs-Duhem relation, which is equivalent to the equation

$$\theta_\alpha \frac{\partial}{\partial\theta_\alpha}\log\lambda_\alpha = \theta_\beta \frac{\partial}{\partial\theta_\beta}\log\lambda_\beta, \tag{16}$$

or
$$\theta_\alpha \frac{\partial}{\partial\theta_\alpha}\log\xi_\alpha = \theta_\beta \frac{\partial}{\partial\theta_\beta}\log\xi_\beta, \tag{17}$$

which shows that $\partial\xi_\alpha/\partial\theta_\alpha$ and $\partial\xi_\beta/\partial\theta_\alpha$ both vanish whenever $\partial\xi/\partial\theta_\alpha$ vanishes.

The Gibbs-Duhem relation is usually derived thermodynamically from the experimental fact that vapour pressures are intensive properties of a liquid phase (i.e. do not depend on the actual extent of the liquid phase), supposing of course that this is large enough for surface effects to be neglected. But it can be derived just as well from statistical theory and follows most simply from equations (7). We find

$$\frac{\partial}{\partial N_\beta}\log\lambda_\alpha = \frac{\partial}{\partial N_\alpha}\log\lambda_\beta,$$

which is equivalent to equation (16) provided only that λ_α and λ_β depend on N_α and N_β only through the ratio N_α/N_β, i.e. that λ_α and λ_β are in fact intensive properties of the liquid phase. This is undoubtedly true and requires that

$$\log\sum_X t(rX, N_\alpha, N_\beta)\, e^{-Xw/kT} \sim (N_\alpha+N_\beta)\,(\text{function of } N_\alpha/N_\beta)$$

as we can easily verify when $w = 0$.

Returning to the condition $x \leqslant 1-2/r$, we note that this cannot be satisfied at any temperature if $w<0$; but if $w>0$ then $x \leqslant 1-2/r$ if $T<T_c$, where $T_c = w/rk\log(r/r-2)$. Thus if $w<0$ there are no critical phenomena, and the partial vapour pressure curves are monotonic, but if $w>0$ then, for $T<T_c$, there are values of ξ_α and ξ_β to which correspond three values of θ_α: we have the phenomenon of the separation of the solution into two phases as described in § 2.

If the solution separates into two phases (differing in relative concentration, so that at low temperatures we have a dilute solution of α in β and a dilute solution of β in α) then, for equilibrium, the partial vapour pressures, or ξ_α and ξ_β, must be the same for the two parts. Equation (17) shows that

$$\log\xi_\beta = \text{constant} + \theta_\alpha\log\xi - \int_0^{\theta_\alpha}\log\xi\, d\theta_\alpha,$$

and so two points on a $\log\xi$ versus θ_α curve which represent liquid phases of differing constitution in equilibrium with each other at the temperature T, must lie on the line $\xi=$ constant which cuts off equal areas (above and below it) from the two hooks of this curve. By symmetry this is the line $\log\xi = 0$ (see fig. 1). The curves corresponding to temperatures greater than T_c (the critical-solution-temperature) cut the line $\log\xi = 0$ only once, at $\theta_\alpha = \frac{1}{2}$, but those for temperatures less than the critical-solution-temperature cut it also at two more points, whose co-ordinates depend on the temperature, at equal distances from this.

A solution having the critical constitution, $\theta_\alpha = \frac{1}{2}$, will start to separate into two phases at the critical-solution-temperature T_c. But one having an initial constitution, θ_α^*, differing from $\frac{1}{2}$, will commence to separate only at a lower temperature, the temperature, T^*, corresponding to the particular $\log \xi$ curve which cuts the axis at $\theta_\alpha = \theta_\alpha^*$. Fig. 2 shows how this temperature depends on the initial constitution. It is interesting that plotting T^*/T_c against θ_α^* gives a curve which should be the same for all such strictly regular solutions with the same co-ordination number. The full curve in fig. 2 is drawn for $r = 8$, while the broken one is found on taking $r = 12$ (putting $r = 12$ in the equations above): the curves are not very different. But we do not expect the above equations to be as accurate in the case of close packing ($r = 12$) as in the case, say, of body-centred cubic packing ($r = 8$) and we shall now briefly consider a more accurate theory for face-centred cubic packing.

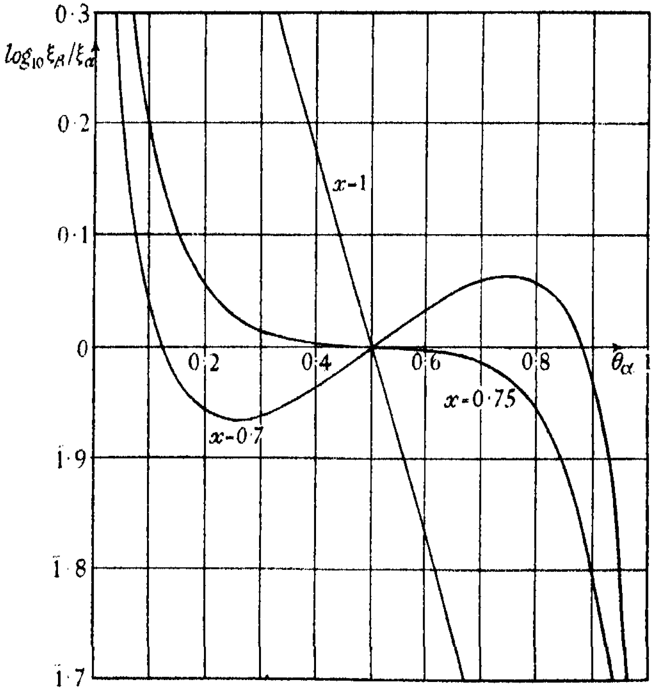


FIG. 1. Curves show the nature of $\log_{10} \xi_\beta/\xi_\alpha$ as a function of θ_α. They are drawn for $r = 8$, with $x = 0{\cdot}7$, $0{\cdot}75$ and 1.

4. FACE-CENTRED CUBIC PACKING

If r is put equal to 12 in the formulae of § 3 the equation determining the critical-solution-temperature becomes

$$x = e^{-w/rkT} = 0{\cdot}833.$$

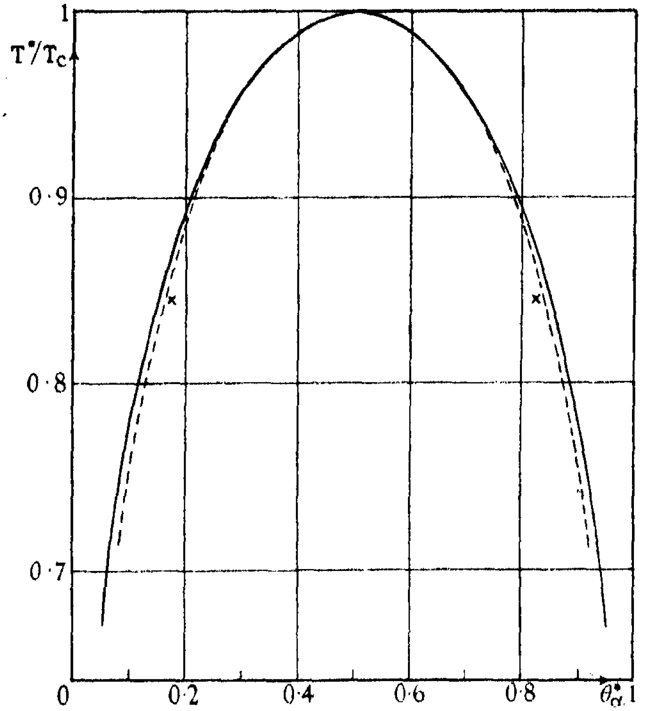


FIG. 2. The full curve is drawn for $r = 8$, the broken one is for $r = 12$, and the crosses mark points calculated on the more exact theory for close packing of § 4.

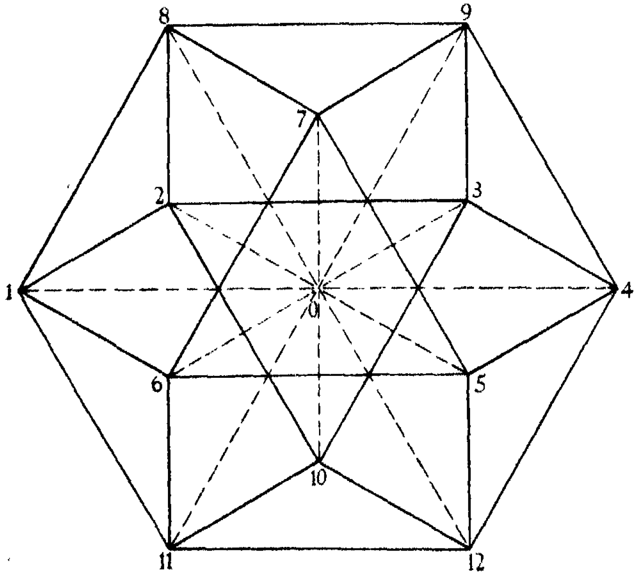


FIG. 3. The geometry of a central site and its first shell for face-centred cubic packing.

We now intend primarily to find to what extent (if at all) it is necessary to modify this result when the interactions between neighbouring first shell molecules are introduced explicitly.

The geometry of a central site and its first shell for face-centred cubic packing is illustrated in fig. 3. O there represents the central site, the points 1, 2, 3, ..., 12 represent its neighbours and the lines show which of these are neighbours of each other—and thus which interactions we are going to consider explicitly.

In forming the statistical weight it is sufficient to use the notation of equations (13′) and (14′): we thus include factors ξ_α, ξ_β and x for each α-molecule, β-molecule and $\alpha\beta$-pair of neighbours, respectively, there being also a factor ϵ for each β-molecule in the first shell. A factor x, however, is necessary, not only when a first shell molecule differs from that in the central site but also when two neighbouring first-shell molecules differ from each other. Thus, for instance, when there are α-molecules at 0, 2, 3, 10, 12 and β-molecules in the other places there is a term $\xi_\alpha^5 \xi_\beta^8 x^{16} \epsilon^8$ and, moreover, there are 47 other configurations exactly equivalent to this one. By adding such terms we find:

The statistical weight of the set of all configurations in which the central site is occupied by an α-molecule is

$$\begin{aligned} Q_\alpha^* = \xi_\alpha[\xi_\alpha^{12} &+ (x\epsilon\xi_\beta)^{12} + (x\epsilon\xi_\beta \xi_\alpha^{11} + (x\epsilon\xi_\beta)^{11}\xi_\alpha)(12x^4) \\ &+ ((x\epsilon\xi_\beta)^2 \xi_\alpha^{10} + (x\epsilon\xi_\beta)^{10}\xi_\alpha^2)(42x^8 + 24x^6) \\ &+ ((x\epsilon\xi_\beta)^3 \xi_\alpha^9 + (x\epsilon\xi_\beta)^9 \xi_\alpha^3)(44x^{12} + 120x^{10} + 48x^8 + 8x^6) \\ &+ ((x\epsilon\xi_\beta)^4 \xi_\alpha^8 + (x\epsilon\xi_\beta)^8 \xi_\alpha^4)(9x^{16} + 96x^{14} + 240x^{12} + 96x^{10} + 54x^8) \\ &+ ((x\epsilon\xi_\beta)^5 \xi_\alpha^7 + (x\epsilon\xi_\beta)^7 \xi_\alpha^5)(108x^{16} + 264x^{14} + 264x^{12} + 120x^{10} + 36x^8) \\ &+ (x\epsilon\xi_\beta)^6 \xi_\alpha^6(216x^{16} + 240x^{14} + 336x^{12} + 96x^{10} + 36x^8)], \end{aligned}$$

and

The statistical weight of the set of all configurations in which the central site is occupied by a β-molecule is

$$\begin{aligned} Q_\beta^* = \xi_\beta[(x\xi_\alpha)^{12} &+ (\epsilon\xi_\beta)^{12} + (\epsilon\xi_\beta(x\xi_\alpha)^{11} + (\epsilon\xi_\beta)^{11} x\xi_\alpha)(12x^4) \\ &+ ((\epsilon\xi_\beta)^2 (x\xi_\alpha)^{10} + (\epsilon\xi_\beta)^{10}(x\xi_\alpha)^2)(42x^8 + 24x^6) \\ &+ ((\epsilon\xi_\beta)^3 (x\xi_\alpha)^9 + (\epsilon\xi_\beta)^9 (x\xi_\alpha)^3)(44x^{12} + 120x^{10} + 48x^8 + 8x^6) \\ &+ ((\epsilon\xi_\beta)^4 (x\xi_\alpha)^8 + (\epsilon\xi_\beta)^8 (x\xi_\alpha)^4)(9x^{16} + 96x^{14} + 240x^{12} + 96x^{10} + 54x^8) \\ &+ ((\epsilon\xi_\beta)^5 (x\xi_\alpha)^7 + (\epsilon\xi_\beta)^7 (x\xi_\alpha)^5)(108x^{16} + 264x^{14} + 264x^{12} + 120x^{10} + 36x^8) \\ &+ (\epsilon\xi_\beta)^6 (x\xi_\alpha)^6 (216x^{16} + 240x^{14} + 336x^{12} + 96x^{10} + 36x^8)]. \end{aligned}$$

The equation for ϵ is most easily found from the condition

$$\xi_\alpha \frac{\partial}{\partial \xi_\alpha} Q^*_\beta = \xi_\beta \frac{\partial}{\partial \xi_\beta} Q^*_\alpha. \tag{18}$$

As before we have

$$\theta_\alpha = Q^*_\alpha/(Q^*_\alpha + Q^*_\beta); \quad \theta_\beta = Q^*_\beta/(Q^*_\alpha + Q^*_\beta),$$

though now equation (12) is replaced by

$$\bar{X} = (\bar{N}_\alpha + \bar{N}_\beta)\,(1/72)\,x \frac{\partial}{\partial x} \log\,(Q^*_\alpha + Q^*_\beta),$$

the factor in the denominator being 72, since the patch now includes 36 bonds explicitly: but we shall not discuss this further except to remark that it furnishes a useful check on the coefficients and indices in the other equations, since when $w = 0$, and the mixing is random, we must have

$$\bar{X}^2 = (\bar{N}_\alpha - \bar{X})\,(\bar{N}_\beta - \bar{X}):$$

this is found to be so.

The only further discussion undertaken of these equations has been numerical. Having formed equation (18) it is convenient to change to the variables

$$\xi = \xi_\beta/\xi_\alpha; \quad \eta = \epsilon\xi,$$

and then we can solve for ξ in terms of η and x, getting a formula similar to (13″) though the coefficients, in the numerator and denominator, are no longer those of binomial expansions. We can thus, graphically, plot ξ as a function of η for various different values of x. The critical value of x will be that for which $\partial\xi/\partial\eta = 0$ at the point $\eta = 1$, $\xi = 1$. For higher values of x $\partial\xi/\partial\eta$ will not vanish at any point of the ξ, η curve, while for lower values of x $\partial\xi/\partial\eta$ will vanish twice, once on each side of the line $\eta = 1$. Fig. 4 shows ξ, η curves for $x = 0{\cdot}7$, $0{\cdot}8$, $0{\cdot}9$ and $0{\cdot}833$. It is only necessary to have the curves for $\eta > 1$, since changing η into $1/\eta$ merely changes ξ into $1/\xi$. From fig. 4 we see that the critical value of x is slightly less than 0·833, but on the other hand it must be greater than 0·826, since, as the inset shows, for this value of x, $\xi = 1$ when $\eta = 1{\cdot}1$.

We conclude therefore that the more accurate approximation in which the interactions between neighbouring first-shell molecules are included explicitly results in a slightly lower value for the critical value of $x \equiv e^{-w/rkT}$—0·828, say, instead of 0·833. The θ_α, $\log \xi$ curves are changed somewhat; for instance, the curve for $x = 0{\cdot}8$ cuts the θ_α axis at $\theta_\alpha = 0{\cdot}137$ on the simpler theory (in which r is put equal to 12 in the equations of § 3), and at $\theta_\alpha = 0{\cdot}175$

on using the expressions of the present paragraph, but as concerning fig. 2 this is considerably counterbalanced by the slight lowering of the critical temperature. The value of T^*/T_c that we now find to correspond to an initial constitution $\theta_\alpha^* = 0{\cdot}175$ (or $= 0{\cdot}825$) is marked in fig. 2, and indicates a slightly greater variation, with the initial constitution, of the temperature at which separation begins. But, as suggested at the end of § 2, a more elaborate discussion of this or a higher approximation (involving explicitly a larger patch of liquid) is hardly appropriate, since it is unlikely that any liquid mixture will have, in practice, a sufficiently regular structure for the refinements to be significant.

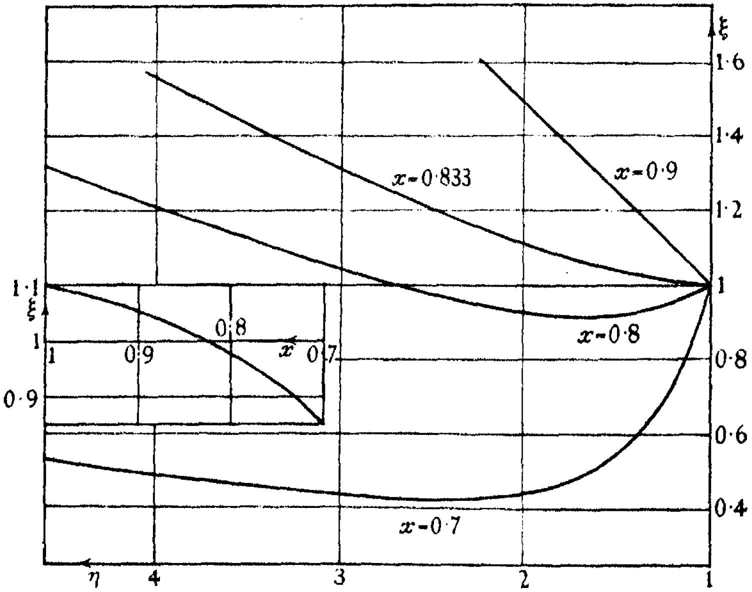


FIG. 4. ξ, η curves for $x = 0{\cdot}7$, $0{\cdot}8$, $0{\cdot}833$ and $0{\cdot}9$. The inset shows, for various x, the values of ξ which correspond to $\eta = 1{\cdot}1$.

In concluding I wish to thank Professor R. H. Fowler and Mr E. A. Guggenheim for helpful criticism and encouragement in the writing of this note, and the Department of Scientific and Industrial Research for a Maintenance Grant.

SUMMARY

An attempt is made to improve and extend Guggenheim's tentative theory of strictly regular binary liquid mixtures by using Bethe's indirect method to avoid the unknown combinatory factor in the partition function for the

liquid phase. The expression found in this way for the configurational energy of the mixture differs somewhat from that proposed by Guggenheim, and consequently formulae for the thermodynamic functions are also modified.

The conditions under which such solutions will exhibit critical mixing phenomena are discussed, and the variation with temperature of the concentration at which the liquid mixture begins to separate into two phases is investigated.

Though for simplicity the theory is developed on the supposition that the molecules of the liquid are not close packed, numerical calculations for face-centred cubic packing, included in the last section, suggest that this does not significantly affect the results.

REFERENCES

Bethe, H. 1935 *Proc. Roy. Soc.* A, **150**, 552–75.
Easthope, C. E. 1937 *Proc. Camb. Phil. Soc.* **33**, 502–17.
Eyring, H. and Hirschfelder, J. 1937 *J. Phys. Chem.* **41**, 249–57, 321–22.
Fowler, R. H. 1936 "Statistical Mechanics", 2nd ed., § 13·4. Camb. Univ. Press.
Guggenheim, E. A. 1935 *Proc. Roy. Soc.* A, **148**, 304–12.
Lennard-Jones, J. E. and Devonshire, A. F. 1937 *Proc. Roy. Soc.* A, **163**, 53–70.

The dielectric polarization of a long chain ketone at constant volume and variable temperature

By Alex Müller

(*Communicated by Sir William Bragg, O.M., P.R.S.—Received* 7 *March* 1938)

Previous work (Müller 1937) on long-chain ketones has shown that the dielectric polarization rises from a low value to a maximum as the temperature of the solid approaches the melting-point. This rise takes place in a temperature interval of about 20° C. The characteristic feature is the rapidly increasing gradient of the polarization/temperature curve. The increase goes parallel with a correspondingly large expansion of the crystal lattice (Müller 1932, 1937), and the question therefore arises to what extent the two phenomena are linked together. The problem dealing with this question is investigated in the present work. The substance, instead of being allowed to expand freely, as it was in the previous investigation, is now enclosed in a steel container where it is kept at constant volume. The effect of this constraint upon the polarization/temperature curve is very striking, as is shown later.

APPARATUS

The electrical equipment is the same as that used before. It consists of a small oscillator emitting waves of about 20 m. The resonating circuit with a variable capacity and the substance condenser is loosely coupled with the oscillator. The new substance condenser has to stand pressures estimated at about 2000 atm.

Considerable difficulties were first encountered in preventing leakage at these pressures. The final successful apparatus was built adopting Bridgeman's technique, which uses the internal pressure to make the joints tight. No trouble occurred with this condenser during a long series of experiments. The design is shown in fig. 1.

The condenser stands on three metal supports in an oil bath which is heated by a Bunsen. The temperature of the bath is read on a mercury thermometer. A second thermometer with a long narrow bulb is placed into a hole in the inner electrode of the condenser, the bulb reaching right down to its centre. The compressibility of the ketone being much larger than that of steel and of the insulating material, corrections due to the deformation of the condenser when under pressure are outside the limit of accuracy aimed at in these experiments. The capacity of the condenser is measured with benzene $\epsilon = 2{\cdot}283$ as a normal, and found to be 1·25 scale divisions of the standard variable condenser or 7·32 cm. absolute units. The total volume of the condenser is approximately 9 c.c.

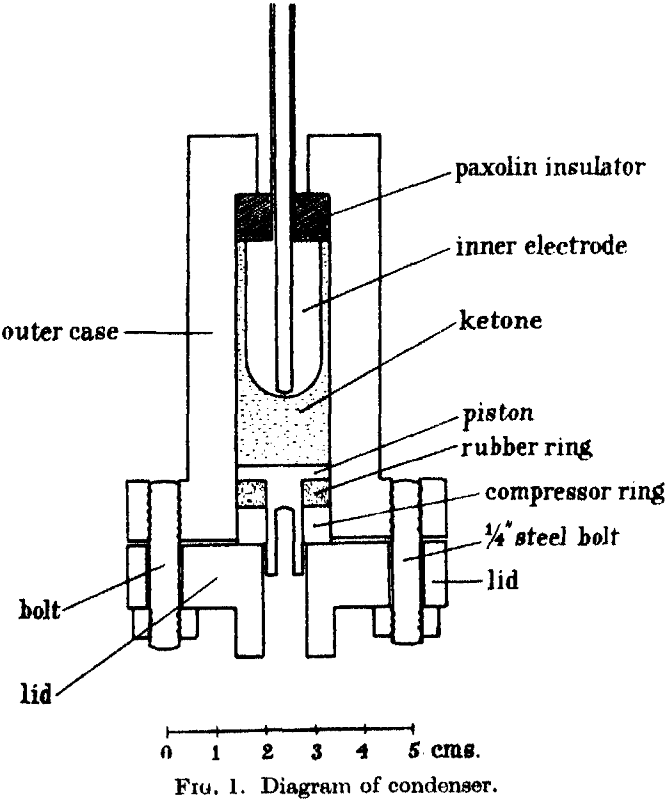


FIG. 1. Diagram of condenser.

MATERIAL

The ketone $CH_3(CH_2)_7\overset{\overset{O}{\|}}{C}(CH_2)_7CH_3$, m.p. 50° C., is identical with one of the substances used in the previous work. The writer is much indebted to Dr J. W. H. Oldham in this laboratory for its synthesis.

RUN OF AN EXPERIMENT

The experiments are carried out in much the same way as before. A few indications may be given for the sake of completeness.

The zero reading with the empty condenser is taken first. The substance is then filled in carefully in small portions so as to avoid air cavities. The piston with the rubber packing is screwed down with the aid of the six bolts in the lid, the initial pressure being adjustable by the amount of tightening of the screw nuts. Capacity variations due to the variation in the position of the closing piston are negligible on account of the comparatively large distance of the piston face from the inner electrode.

The first readings are usually taken at room temperature. After that the temperature of the oil bath is raised in convenient steps. The oil is stirred during the heating process. When both thermometers show the same temperature after removal of the heat source, the readings are taken on the scale of the variable condenser. Observations made in this way usually reproduce within 1–2 % over periods of days. The zero readings behave in the same way.

RESULTS

The dielectric capacity ϵ is calculated from the observations by the following expression:

$$\epsilon = 1 + \frac{r_0 - r_t}{1{\cdot}25},$$

where r_0 is the zero reading with the empty condenser, r_t the reading taken at the temperature t with the condenser filled with the ketone. 1·25 is the capacity of the empty condenser in scale divisions of the variable capacity. Fig. 2 shows a number of typical curves representing $\epsilon - 1/\epsilon + 2$ plotted against the temperature. Curve I represents an experiment in which the condenser is filled at high initial pressure. Owing to this pressure the melting-point moves upwards and becomes so high in this particular case that the substance remains solid even at 80° C. The dotted curve drawn for comparison shows the behaviour of the substance at atmospheric pressure, the melting-point marked by an arrow is at 50° C. The direct test for the non-melting of the substance is done with the aid of a small steel ball put at the bottom of the condenser before filling. The actual position of the condenser during the experiment is inverted. If, as is observed with the

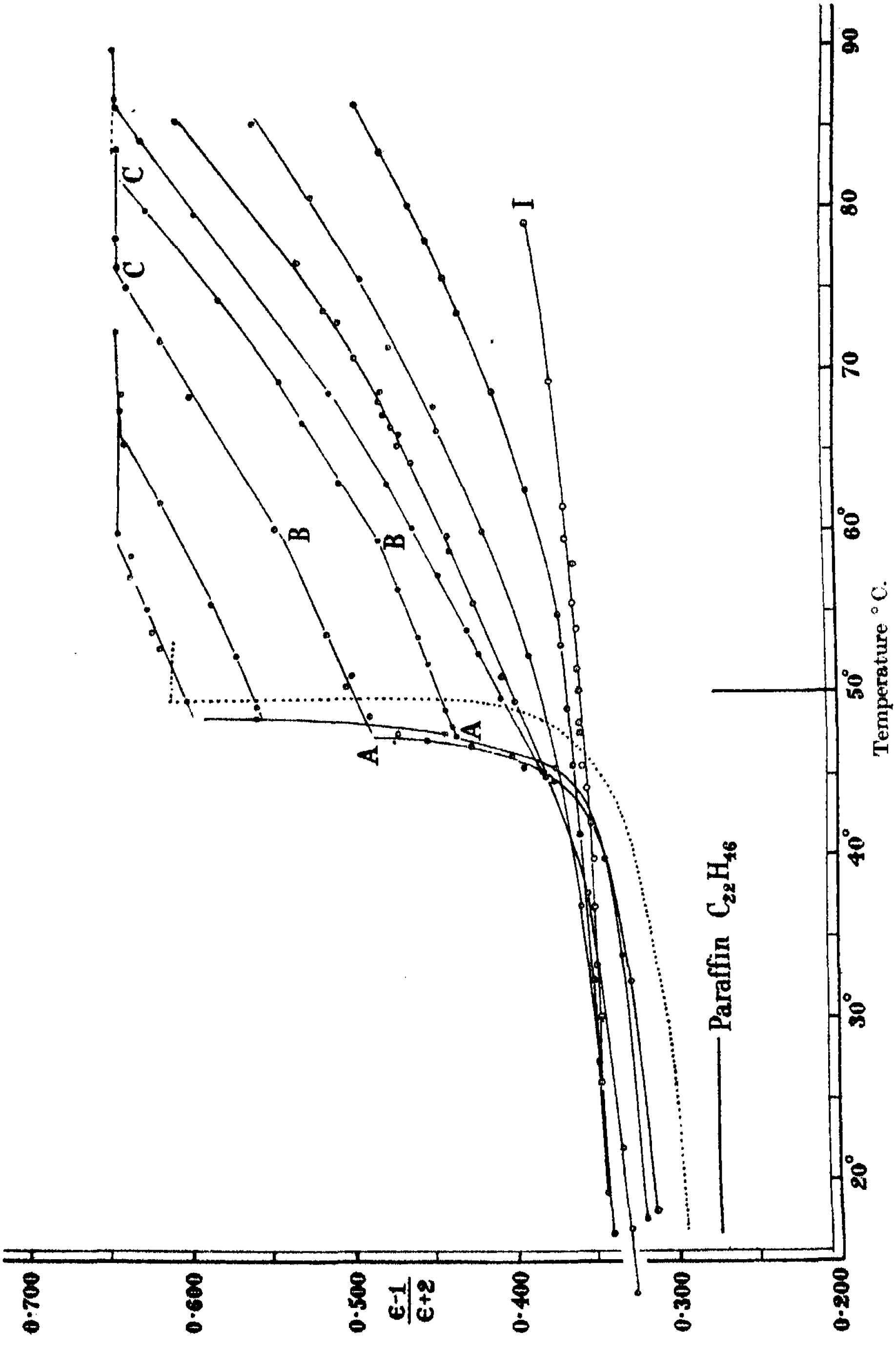


FIG. 2. Observed $(\epsilon-1)/(\epsilon+2)$ of ketone $CH_3(CH_2)_7C_{11}(CH_2)_7CH_3$.

experiments at high pressure, the steel ball remains in its position, it is a sign that the substance did not melt.

The subsequent curves on the diagram are those observed with gradually decreasing initial pressure, the substance being kept at constant volume on each individual curve, until finally a stage is reached where the filling of the condenser is incomplete at first. Here the substance expands first freely on heating and the polarization follows that previously measured at atmospheric pressure. As soon as the pressure sets in the curve shows a sharp kink *A* after which the rise is much slower than before. On heating the substance still further a second kink *B* appears. This marks the temperature at which the first melting sets in. With still increasing temperature more and more of the solid phase disappears until finally the whole of the ketone is molten. At this point *C* the polarization has reached its maximum and decreases slowly with further rise in temperature.

DISCUSSION OF THE RESULTS

The diagrams suggest that there are three distinct temperature ranges, each range indicating the prevalence of certain definite physical conditions. These three ranges show in experiments, at both constant pressure and at constant volume.

In the first and lowest temperature range the polarization is practically independent of the temperature, in the second intermediate range there is a very marked rise, and in the third and highest range the polarization decreases slowly with increasing temperature.

These ranges are to be considered. In the first range the ketone molecules are kept in their positions in the lattice by the Van der Waals and dipole forces. They perform small oscillations due to an elastic restoring force under the influence of the temperature. An external electric field has a twofold effect: it polarizes the molecule by separating the electric charges, and, secondly, polarizes the material by displacement of the dipoles. The first effect, essentially the polarization of the paraffin chain in the ketone, does not interest us here. It is quite unaffected by temperature changes or by changes in the arrangement of the molecules. The dipole effect is different. In the present material the dipole is attached to a long paraffin chain exactly halfway between the ends. Owing to this symmetrical position it is very likely that the line connecting the two charges of the dipole is perpendicular to the chain axis or at least not far from this position. The chain axes are fixed in relation to each other, since we are dealing with a solid, and are evenly distributed over all possible directions in space. The system as a whole has therefore no resulting dipole moment. Such a moment appears, however, when the material is placed in an electric field. The dipoles tend to orient themselves, their most probable movement being a rotation round the fixed chain axes. This rotation is opposed by the elastic forces. The resulting position of equilibrium determines the magnitude of the induced polarization, which is found to be equal to

$$\frac{\mu^2}{3W}.$$

μ is the dipole moment. The potential is assumed to be $\frac{1}{2}W\sin^2\alpha$, i.e. it has for small displacements the conventional character of an elastic potential. For large angles it suggests the necessarily periodic nature of the field of force to which the dipole is subjected in the present model. So far the effect of the temperature is neglected. Now it is easily seen from first principles that the temperature becomes only effective when kT is comparable with W. For any value of kT below this threshold the polarization remains constant. This is shown explicitly in the calculations of R. H. Fowler (1935), and explains the first range. Before discussing the second range, the most important in the present work, a few words may be said about the third. Here the substance is completely molten, and its behaviour is accounted for by the classical theory of Debye. So far there is hardly any difference between the experiments made at constant pressure and those at constant volume. The situation changes completely in the second range. The rise of the polarization when observed at constant pressure takes place in a temperature interval of about 20° C. below the melting-point. This interval increases considerably when the experiment is made at constant volume. One reason for this is obvious. The melting, instead of taking place at one definite temperature, is spread over an interval, when the substance is under constraint. Further, as the experiments show, there is a considerable interval preceding the melting in which the polarization rises from its lowest level. The position of the second range on the temperature scale, and the shape of the polarization temperature curves, depend upon the initial pressure at which the substance is filled in. Take curve I on fig. 2. In the corresponding experiment the initial pressure is so high that the substance remains solid up to the highest temperature shown on the drawing. The interval of rising polarization is very large. It decreases as the initial pressure gets lower.

It is particularly interesting to compare the constant-pressure and constant-volume curves in the region below the melting-point. The constant-pressure curve rises rapidly with increasing gradient as the temperature approaches the melting-point, whereas the constant-volume curve is nearly a straight line with a comparatively small slope.

It has been pointed out in the introduction that crystals consisting of chain molecules have a particularly large thermal expansion in the neighbourhood of their melting-point. The present experiments show that the rapid change of the polarization which simultaneously takes place in this temperature range becomes much smaller when the expansion is suppressed. This suggests that the volume change is primarily responsible for the increased mobility of the molecules.

Some time ago I made an approximate calculation of the lattice energy of paraffin crystals (Müller 1936*a*), showing that this energy is essentially accounted for by the Van der Waals forces. In a continuation of this work (Müller 1936*b*) an estimate was obtained for the threshold energy W. The numerical value of W for one CH_2 group is found to be approximately 5×10^{-14} ergs. For a molecule with say $20CH_2$ groups it is twenty times as much, i.e. 1×10^{-12} ergs. This value holds for a crystal at room temperature before the large expansion has set in. The average thermal energy of a free rotator with a fixed axis at this temperature is $\frac{kT}{2} = 2 \times 10^{-14}$ ergs, which is much smaller than the threshold value. Rotation is therefore extremely infrequent. After the expansion, that is to say when the crystal is at the melting-point, the threshold value drops to the same level with 2×10^{-14} ergs, as the calculation shows and rotation is to be expected. This suggests for the present experiments that the molecules in the expanded crystal are free to orient themselves in the external field which acts on the attached dipoles.

The argument which makes the expansion alone responsible for the increase of the polarization at constant pressure is obviously incomplete. It does not take into account the increase existing when the system is kept at constant volume. The fact that it provides an approximate explanation simply shows that the expansion plays an important role.

A theoretical discussion of the present observations is outside the scope of this paper and is reserved for a future publication. There is, however, one observation which seems to simplify matters, which will be discussed now.

The transition B between the one- and the two-phase system at the beginning of the melting is not very prominent at low pressures and becomes practically undetectable at high pressures. This indicates that the two systems cannot differ much in their internal arrangement. We take first the one-phase system at low temperature. The molecules perform oscillations relative to a fixed position of equilibrium. The external field producing only a slight perturbation has very little effect so long as the elastic restoring forces are present. These forces diminish and the amplitudes of oscillation become larger, but so long as the oscillation centres persist the polarization remains small. Finally, when the energy due to the external field has the same order of magnitude as the remaining threshold energy the dipoles will follow the external field, but since the energy of the field is very small the molecules must be practically free. We therefore conclude that the increase of the polarization observed with the solid is due to an increase in the number of free molecules. Since we have every reason to believe that free rotation exists in the liquid phase we come to the conclusion that the solid and the liquid cannot differ very much in the region where melting begins. This conclusion is in accordance with observation. A difference between the two phases must necessarily remain. It is magnified at the upper limit of the melting process where the discontinuity C persists. This is not surprising. The observed physical property, the polarization in the present experiments, is much larger at the upper end of the curve, and any discontinuity is bound to be enhanced more or less in proportion to the magnitude of the observed property.

The argument in the previous section leads to the conclusion that the system consists essentially of a mixture of rotating and non-rotating dipoles.

The discussion of the numerical data which involves a statistical treatment of this dipole system is to be given later. The aim of the present paper is to prepare the ground for an investigation of the melting of chain compounds.

In conclusion the writer thanks the Director and the Managers of the Royal Institution for their kind interest in the work.

SUMMARY

The present work is a continuation of a previous investigation on the dielectric polarization of a long-chain ketone. The material is again a chain molecule carrying a dipole at its centre. The substance, instead of being allowed to expand freely, as in the previous experiments, is kept at constant volume, and the effect of temperature variation is now studied under these new conditions.

The experiments show that the constraint imposed upon the system has a very marked effect on the dielectric polarization. It follows from the

observations that the rapid rise of the polarization observed in the previous constant-pressure experiments is intimately connected with the similarly large expansion of the crystal lattice found with all long-chain substances.

With regard to the theory of the co-operative effects shown by these substances, the present experiments prove that a variation of the boundary conditions has a large effect upon the behaviour of the substance.

References

Fowler, R. H. 1935 *Proc. Roy. Soc.* A, **149**, 1.
Müller, A. 1932 *Proc. Roy. Soc.* A, **138**, 514.
— 1936*a* *Proc. Roy. Soc.* A, **154**, 624.
— 1936*b* *Helv. Phys. Acta*, **9**, 626.
— 1937 *Proc. Roy. Soc.* A, **158**, 403, 414.

Diamagnetism of cadmium

By S. Ramachandra Rao and S. Sriraman
Annamalai University, Annamalainagar, South India

(*Communicated by O. W. Richardson, F.R.S.—Received* 18 *January* 1938)

1. Introduction

The study of the susceptibility of metal crystals has received some attention in recent years both from theoretical and experimental points of view. The careful and thorough investigations of Goetz and Focke (1934) on single crystals of bismuth have greatly extended our knowledge regarding the properties of the crystalline state. McLennan, Ruedy and Cohen (1928) and McLennan and Cohen (1929) have determined the principal susceptibilities of a number of metals, including zinc, cadmium, antimony and tin. Hoge (1935) investigated tin crystals and showed the existence of a feeble anisotropy. Rao and Subramaniam (1936) studied thallium and showed from magnetic data the existence of a transformation point (235° C.), at which temperature its ordinary hexagonal structure passes over into the cubic type. Rao (1936*a*, *b*) has examined the influence of cold-working on single crystals of bismuth, zinc and tin and arrived at important conclusions based on the observed variations of the principal susceptibilities. An extension of this study to other metals and a careful repetition of some of the earlier investigations on the lines of more recent work seem essential. In this paper cadmium has been taken up for special study. We have indeed the values of McLennan and others (1928) for the principal susceptibilities of cadmium, but it is proposed here to study carefully the influence of cold-working and of foreign matter on its diamagnetism.

On the theoretical side a large amount of interesting work has been accomplished. Pauli (1926) applied the Fermi-Dirac statistics to the calculation of the paramagnetism of a free electron gas. If $(\chi_A)_e$ is the gram-atomic susceptibility due to the free electrons, q the number of free electrons per atom, and V_0 the width of the occupied energy range in the completely degenerate state in volts, for ordinary temperatures,

$$(\chi_A)_e \times 10^6 = 48\cdot17\left(\frac{q}{V_0}\right)\left\{1-6\cdot11\times10^{-9}\left(\frac{T}{V_0}\right)^2\right\}. \tag{1}$$

Landau (1930) has shown that for strictly free electrons, there is also a diamagnetic effect arising from the discreteness of the energy states in a magnetic field. This effect is found to be numerically a third of the paramagnetic effect. Neglecting the small term proportional to $(T/V_0)^2$ in the above expression, we have, for the gram-atomic susceptibility of free electrons,

$$(\chi_A)_e \times 10^6 = 32{\cdot}11q/V_0. \tag{2}$$

Peierls (1933) has discussed the diamagnetism of quasi-bound electrons. He shows that the condition for the applicability of Landau's formula to metals is given by $kT \gg h\tau$, where τ is the average time between collisions. By considerations of the entropy of the electron gas, Landau (1930) has arrived at the following expression for the diamagnetism of free electrons

$$\chi_L = 0{\cdot}623 n_0^{\frac{1}{3}} \rho^{-\frac{2}{3}} A^{-\frac{1}{3}} \times 10^{-6}, \tag{3}$$

where n_0 is the number of free electrons per atom in the metal, ρ is the density and A is the atomic weight. According to this calculation, the total paramagnetic susceptibility of the degenerate electron gas is $2\chi_L$, since the paramagnetic effect due to electron spin is three times as great as the Landau diamagnetism. In crystals we have the periodic field due to the ions in the lattice. Peierls (1933) has studied the effect of the lattice field on the diamagnetism of metals. The applications to the case of cadmium will be taken up later in the paper.

2. Experiment

(a) *Metal*

The cadmium metal used in this investigation was of the purest quality from Kahlbaum. A spectroscopic analysis of the metal showed the presence of 0·01 % of zinc and 0·001 % of lead. The specimens were tested for ferromagnetic impurities by investigating the susceptibility of the cast cadmium rods at different field strengths. The details will be described presently. The metal was found to have a density of 8·626.

(b) *Preparation of polycrystalline rods*

For the magnetic investigations, polycrystalline rods had to be prepared of length about 10 cm. and diameters between 2·5 and 5 mm. Diameters of this order were found to be specially convenient for the magnetic work. The casts were made with the help of tubes similar to the one shown in fig. 1. After cleaning the tube thoroughly, a sufficient quantity of the metal was taken in the tube D. The end C was then connected to a high vacuum outfit consisting of a Cenco hyvac pump and a mercury diffusion pump and the pressure reduced to the order of 10^{-5} mm. The metal in *D* was next melted, and after all the absorbed gases were removed the clean metal was allowed to flow in *AB*.

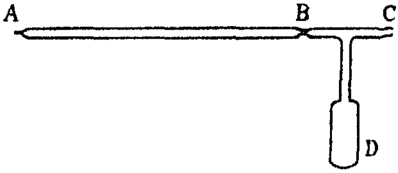


FIG. 1

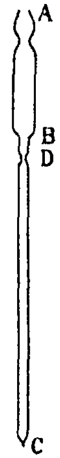


FIG. 2

In later investigations, a tube of the kind shown in fig. 2 was used. *AB* is a tube of 1 cm. diameter with the tube *BC* joined at *B*. Pure metal bits were placed in *AB* and the tube was joined at *A* to the high vacuum outfit. With all-glass connexions it was possible to reach the sticking pressure in a Macleod gauge connected to the system. When such low pressures were reached the metal was heated at *B* with the pumps running. When the pressure read in the gauge was of the order 10^{-5} mm., the metal at *B* was heated strongly and melted. When the tube *BC* was almost half full, the tube was sealed off at *D*. For investigations on polycrystalline metal, the cast rods were removed by cracking the glass all around and breaking.

(c) *Preparation of single crystals*

Single crystals were prepared by the method of Bridgman (1925). The tube containing the metal prepared in the foregoing section was suspended vertically along the axis of an electrically heated furnace, the temperature inside being regulated to be about 350° C., nearly 30° above the melting-point of cadmium. The lowering device consisted of an ordinary time piece, the setting screw of which was replaced by a cylinder. The circumference of this cylinder was adjusted to give a rate of lowering of 3 cm./hr. for the metal used. As the metal tube emerged from the heater, the lower end was gradually cooled. The steep temperature gradient set up inside brought about a gradual crystallization. Cadmium crystals thus prepared were found to be good specimens of single crystals.

(d) *Crystal structure*

Cadmium crystallizes in the hexagonal close-packed system, the *a* and *c* values (in Angstrom units) according to Clark (1932) being 2·973 and 5·606 and ($c/a = 1{\cdot}885$). The basal plane is the 0001 plane which is also the cleavage plane when the crystal is broken or etched. The crystals could be etched with dilute hydrochloric acid or when sufficiently thin could be broken after cooling with carbon dioxide snow and ether. The angle ϕ made by the hexagonal axis with the cylindrical axis was determined for each crystal by using a spectrometer after the magnetic investigations were over.

(e) *Determination of the susceptibility of the polycrystalline specimens*

The Gouy method was adopted for the measurement of the susceptibility of the metal specimens. The electromagnet was one of the Pye type, and the pole faces of diameter 3 cm. were fixed parallel to each other and at a distance of 1½ cm. For any given sample, measurements were made at different field strengths by varying the current through the electromagnet from 2 to 7 amp. The intensities of the magnetic field varied under these conditions from 8·2 to 18·8 k.gauss. The cylindrical rod of metal was suspended by a fine wire from one arm of a chemical balance. The specimen was arranged to have its axis vertical and the lower metal surface to lie symmetrically between the pole faces. The weight of the rods were determined first in the absence of the magnetic field and then successively for different currents through the electromagnet. A final reading was taken again in the absence of the magnetic field. Water was used as the standard substance. A glass tube of uniform diameter and about 20 cm. in length was half-filled with water and the Gouy forces determined at different field strengths on the water column.

If k_m is the volume susceptibility of the metal specimen, a_m its area of section and H the intensity of the field between the pole faces, then the Gouy force is given by

$$mg = \tfrac{1}{2}k_m a_m (H^2 - H_m^2).$$

For water

$$m'g = \tfrac{1}{2}k_w a_w (H^2 - H_w^2).$$

In these expressions H_m and H_w are the intensities of the magnetic field at the other ends of the metal rod and water column respectively. Hence

$$k_m = k_w \frac{a_w}{a_m} \cdot \frac{H^2 - H_w^2}{H^2 - H_m^2} \cdot \frac{m}{m'}$$

or

$$\chi = \frac{k_m}{\rho} = \frac{k_w}{\rho} \cdot \frac{a_w}{a_m} \cdot \frac{H^2 - H_w^2}{H^2 - H_m^2} \cdot \frac{m}{m'}.$$

Here ρ is the density of the metal.

If we now apply a correction for the volume susceptibility k_a for air,

$$\chi = \frac{k_m}{\rho} = \frac{k_a}{\rho} + \frac{k_w - k_a}{\rho} \cdot \frac{d_w^2}{d_m^2} \cdot \frac{H^2 - H_w^2}{H^2 - H_m^2} \cdot \frac{m}{m'},$$

where d_w and d_m are the diameters of the water column and the metal rod. If $d_w = d_m$, the lengths used being cut from the same tubing and if H_w^2 and H_m^2 are negligible when compared with H^2,

$$\chi = \frac{k_m}{\rho} = \frac{k_a}{\rho} + \frac{k_w - k_a}{\rho} \cdot \frac{m}{m'},$$

k_a was assumed to be 0·029* (paramagnetic) at 30° C. following the data of Sone (1920).

(*f*) *Determination of the principal susceptibilities of single crystals of cadmium*

The crystal prepared by the Bridgman method was removed from the glass and suspended from the right pan of the chemical balance. A circular paper with its circumference graduated in degrees was pasted on the pan through the centre of which passed a thick wire carrying the crystal at the lower end. A pointer on the disk gave the reading corresponding to any position of the crystal. Great care was taken to see that the crystal was vertical and that its lower end was symmetrically between the pole faces.

The Gouy forces for any crystal were measured at different positions of the pointer on the circular scale. Measurements were usually taken at intervals of 15° during one complete rotation of the crystal, the current in the electromagnet being kept steady at 5 amp. It was found that when the Gouy forces were plotted against the readings on the circular scale, there were maxima and minima of diamagnetic susceptibility alternately occurring at intervals of 90°. The crystal was successively placed in the position of maximum and minimum Gouy force and, in each case, the susceptibilities were determined for different field strengths. On drawing a graph between the susceptibility and $1/H$, a straight line is obtained which when produced gives the value when $1/H$ is zero, i.e. when H is infinite. This point is apparent from the following relation of Honda (1910):

$$\chi_t = \chi_p + \frac{\sigma m}{H},$$

where χ_t and χ_p are specific susceptibilities of the impure and pure metals, σ the specific intensity of magnetization and m the mass of ferromagnetic impurity present in unit mass of the metal. Here m is assumed to be very small. The following theory will explain how the principal susceptibilities of the crystal were calculated.

---

* All susceptibility values in this paper are diamagnetic (unless otherwise stated) and in 10^{-6} units.

(g) *Theory of the measurements*

Following Jackson (1926), Rao (1936*b*) has given the theory of the determination of the principal magnetic susceptibilities, for a hexagonal crystal. If χ_a, χ_b and χ_c are the principal susceptibilities, the susceptibility χ_θ, in any direction of H making direction cosines α, β, γ with directions of the principal susceptibilities of the crystal, is given by

$$\chi_\theta = \chi_a\alpha^2 + \chi_b\beta^2 + \chi_c\gamma^2.$$

If ϕ is the angle between the cylindrical axis and the c axis of the crystal (fig. 3), we obtain, remembering that $\chi_a = \chi_b$,

$$\chi_\theta = \chi_a(\cos^2\phi\cos^2\theta + \sin^2\theta) + \chi_c\sin^2\phi\cos^2\theta.$$

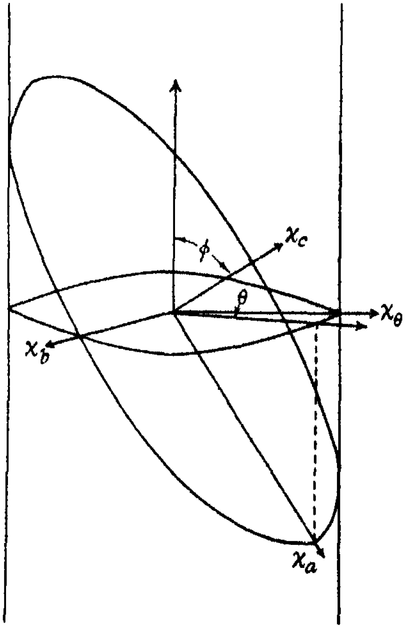


FIG. 3

If χ_0 and χ_{90} are the values when $\theta = 0$ and when $\theta = 90°$ respectively, we have

$$\chi_{90} = \chi_a \quad \text{and} \quad \chi_0 = \chi_a\cos^2\phi + \chi_c\sin^2\phi.$$

If we now designate χ_a and χ_c as $\chi_\perp$ and $\chi_\parallel$, being the susceptibilities perpendicular and parallel to the hexagonal axis,

$$\chi_\perp = \chi_{90} \text{ and } \chi_\parallel = \frac{\chi_0 - \chi_{90}\cos^2\phi}{\sin^2\phi}.$$

From these expressions we can write for χ_θ

$$\chi_\theta = \chi_0 + (\chi_{90} - \chi_0)\sin^2\theta,$$

showing how χ_θ varies with θ. This variation follows the $\sin^2\theta$ graph, the maximum and minimum values being χ_{90} and χ_0 respectively.

(*h*) *Cold-working*

The influence of cold-working on the susceptibilities of polycrystalline cadmium and of its single crystals, was also investigated. The rods were subjected to a tensional stress by the application of weights. The ends were then cleaned with dilute hydrochloric acid. Every care was taken to prevent the metals from coming into direct contact with iron. The densities of the unstrained and strained specimens were determined by the usual hydrostatic method of weighing in air and in water.

(*i*) *Preparation of alloys*

The influence of small quantities of foreign matter was also investigated. Alloys of cadmium containing small quantities of zinc and lead were prepared by the method explained in § 2 (*b*). Weighed quantities of cadmium and the other metal (zinc or lead) were placed in *D* and melted in vacuum. After shaking the tube to get the molten metals mixed uniformly, the alloy was made to run down the tube *AB* to the desired length. The crystallization of the alloy was carried out in the same manner as for the pure metal.

3. RESULTS

(*a*) *Polycrystalline metal*

The susceptibility of the pure polycrystalline metal was determined by measuring the Gouy force on a few specimen rods at different field strengths. The results obtained with a typical specimen of diameter 0·411 cm. and length 9·6 cm. are given in Table I below.

On drawing the straight line graph between $1/H$ and χ and extrapolating it, we obtain for the susceptibility at infinite field strength the value of 0·1740. The mean of a number of measurements with half a dozen rods gave the value of 0·1746 for the diamagnetic susceptibility of cadmium.

The results obtained by other investigators are shown in Table II for comparison.

TABLE I

Field current in amp.	H in k.gauss	$1/H \times 10$	Gouy force in mg.	χ
2	8·2	1·220	6·7	0·1720
3	13·9	0·719	18·9	0·1732
4	16·0	0·625	25·2	0·1728
5	17·3	0·578	29·5	0·1734
6	18·2	0·549	32·7	0·1734
7	18·8	0·532	34·9	0·1735

TABLE II

Investigators	χ
Honda (1910)	0·169
Owen (1912)	0·185
McLennan and others (1928)*	0·194
Meara (1932)	0·180
Rao (1936*a*)	0·174
Authors	0·1746

* Deduced from the principal susceptibilities.

It will be noticed that our value agrees satisfactorily with those of other investigators.

(*b*) *Cold-working on polycrystalline metal*

The effect of cold-working on polycrystalline rods of cadmium has been studied by Rao (1936*a*) who arrived at the conclusion that, due to cold-working, very small changes were produced on the susceptibility of the metal. Table III gives the values obtained by us for some cold-worked specimens.

TABLE III

Relative density of the specimen	Decrease of density due to cold-working	χ
8·626	0·000	0·1745
8·602	0·024	0·1742
8·575	0·051	0·1735
8·530	0·096	0·1730
8·496	0·130	0·1710

Fig. 4 shows the graph between $\delta\rho$ and χ. It will be seen that the diamagnetic susceptibility shows a small decrease on cold-working and that the variation is fairly well represented by a straight line. Points obtained by us are indicated by dark spots while those of Rao (1936*a*) are shown by hollow circles.

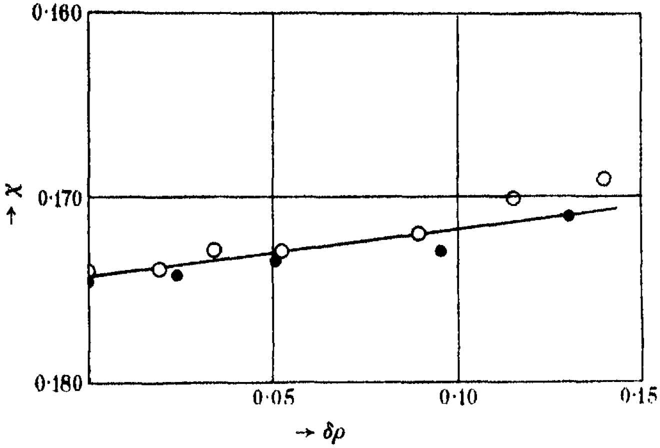


FIG. 4

(*c*) *Principal susceptibilities of the cadmium crystal*

In Table IV are given the values of the Gouy force for various angles of θ ranging from 0° to 360° in the case of a cadmium single crystal of length 9·1 cm. and diameter 4·19 mm. The current through the electromagnet was kept at 5 amp. during these measurements.

TABLE IV

θ	Gouy force in mg. wt.	θ	Gouy force in mg. wt.	θ	Gouy force in mg. wt.
0	32·4	135	38·0	240	30·4
15	31·3	140	38·2	255	31·0
30	30·5	145	38·0	270	32·5
45	30·2	150	37·0	285	34·3
50	30·0	165	34·8	300	36·2
55	30·1	180	32·5	315	38·1
60	30·3	195	31·4	320	38·2
75	31·0	210	30·6	325	38·1
90	32·6	225	30·1	330	37·2
105	34·1	230	30·0	345	35·4
120	36·2	235	30·1	360	32·4

In fig. 5 is shown the graph between θ and the Gouy force on the specimen. The maximum force occurs when $\theta = 140°$ or 320° and the minimum when $\theta = 50°$ or 230°. The crystal was next placed in these positions and the

Gouy forces were determined in each position for different field strengths ranging from 8·2 to 18·8 k.gauss. The results are given in Table V.

TABLE V

Field current in amp.	$\theta = 50°$		$\theta = 140°$	
	Gouy force in mg. wt.	χ	Gouy force in mg. wt.	χ
2	6·6	0·1620	8·5	0·2080
3	19·4	0·1634	24·7	0·2080
4	25·7	0·1636	32·7	0·2082
5	30·0	0·1635	38·2	0·2082
6	32·9	0·1633	42·0	0·2084
7	35·4	0·1635	45·1	0·2083

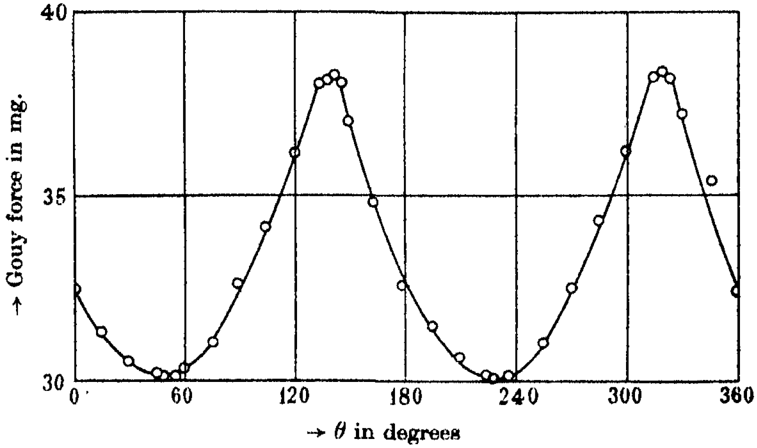


FIG. 5

On extrapolating the χ, $1/H$ graphs to infinite field strengths, we obtain for the susceptibility in the positions of maximum and minimum Gouy force (say χ_0 and χ_{90}) the values of 0·209 and 0·164 respectively. The angle ϕ for the crystal was found by measurement to be 63° 30′. Since $\chi_\perp = \chi_{90}$ and $\chi_\parallel = \frac{\chi_0 - \chi_{90}\cos^2\phi}{\sin^2\phi}$ we obtain the following values:

$$\chi_\perp = 0{\cdot}164, \quad \chi_\parallel = 0{\cdot}220.$$

In this manner the principal susceptibilities of cadmium were determined for five crystals and the values are tabulated below (Table VI).

We thus obtain the following mean values for the principal susceptibilities of cadmium:

$$\chi_\perp = 0{\cdot}163, \quad \chi_\parallel = 0{\cdot}223.$$

TABLE VI

Crystal no.	ϕ	χ_{90}	χ_0	$\chi_\perp$	$\chi_\parallel$
1	63° 30′	0·164	0·209	0·164	0·220
2	62° 30′	0·162	0·212	0·162	0·225
3	62° 30′	0·163	0·210	0·163	0·223
4	62° 0′	0·165	0·210	0·165	0·223
5	61° 0′	0·164	0·211	0·164	0·225
			Mean	0·163	0·223

The mean susceptibility of an ideal polycrystalline aggregate of cadmium arranged at random is $\frac{1}{3}(\chi_\parallel + 2\chi_\perp) = 0{\cdot}183$. McLennan and others (1928) have given the following values:

$$\chi_\perp = 0{\cdot}160, \quad \chi_\parallel = 0{\cdot}261; \quad \text{Mean } \chi = 0{\cdot}194.$$

On comparing these values we find that while there is agreement of the $\chi_\perp$ values, our value of $\chi_\parallel$ is smaller than that of McLennan and others (1928). The mean susceptibility obtained in this investigation is smaller than that of these workers but greater than the value (0·1746) for the polycrystalline aggregate obtained directly above. Attention may be drawn to the fact that a similar discrepancy was obtained by Rao (1936*b*) for zinc. The mean susceptibility of this metal calculated from the principal susceptibilities was 0·167 while the value determined directly was 0·160.

The magnetic results for zinc and cadmium crystals are given in Table VII.

TABLE VII

	Zinc			Cadmium		
Investigators	$\chi_\perp$	$\chi_\parallel$	$\chi_\parallel/\chi_\perp$	$\chi_\perp$	$\chi_\parallel$	$\chi_\parallel/\chi_\perp$
McLennan and others (1928)	0·145	0·190	1·310	0·160	0·261	1·631
Rao (1936*b*)	0·149	0·202	1·356	—	—	—
Authors	—	—	—	0·163	0·223	1·368

It will be seen that the magnetic anisotropy of cadmium obtained here is nearer the value for zinc than the measurements of McLennan and others (1928).

(*d*) *Cold-working of the single crystals*

The crystalline properties of cadmium are similar to those of zinc. On increasing the tensional force applied to one of these crystals, slip bands appeared at stresses of about 0·4 kg. per sq. mm. The tension was increased to five or six times this value. The extensions were measured with a travelling microscope by finding the distances between two marks made on the crystal before and after the cold-working. As the following table will show, the density of the crystal did not vary with cold-working. After the application of tensional stress the principal susceptibilities were determined as outlined in the foregoing section. The following table gives the variations of the principal susceptibilities as the cadmium crystals were subjected to cold-working.

TABLE VIII

$\delta L/L$	Relative density	$\chi_\perp$	$\chi_\parallel$	$\chi_\parallel/\chi_\perp$	Mean χ
0·000	8·640	0·164	0·225	1·372	0·184
0·008	8·638	0·162	0·224	1·383	0·183
0·015	8·638	0·163	0·224	1·374	0·183

An examination of Table VIII shows that the principal susceptibilities alter very little on cold-working. The magnetic anisotropy and the mean susceptibility in the last column also do not show any change. It is therefore most likely that the small changes of susceptibility obtained when polycrystalline rods of cadmium were subjected to cold-working are due to the changes of density accompanying such cold-working.

These results show that the stress effects that may arise during the process of solidification of the metal in the glass tube would not appreciably affect the susceptibility values.

(e) *Influence of small amounts of lead*

Lead in small quantities is known to dissolve in cadmium. In fact it is well known that lead is one of the common impurities in cadmium. It appeared therefore important to study the principal susceptibilities of cadmium crystals containing small quantities of lead. Goetz and Focke (1934) showed that the magnetic anisotropy of bismuth was greatly affected by the nature of the foreign element, elements towards the alkali side of the periodic table increasing the anisotropy and those on the halogen side decreasing this ratio. Similar investigations were conducted by Hoge (1935) on tin.

The lead used was of extra-pure quality from Kahlbaum. Rao and Subramaniam (1936), working with the sample used, obtained for it a diamagnetic susceptibility of 0·107.

Known masses of cadmium and lead to give any desired composition were introduced into the bulb *D* in fig. 1 and single crystals were prepared as detailed in § 2 (*i*) above.

The densities of the specimens were also determined accurately by the hydrostatic method. The results are given in Table IX.

TABLE IX

Specimen no.	Percentage by weight of lead	Relative density	$\chi_\perp$	$\chi_\parallel$	$\chi_\parallel/\chi_\perp$	Mean χ
1	0	8·640	0·163	0·223	1·368	0·183
2	0·37	8·649	0·161	0·224	1·391	0·182
3	0·82	8·660	0·158	0·222	1·405	0·179
4	0·93	8·662	0·157	0·222	1·414	0·179

A scrutiny of Table IX will show definitely that the influence of lead is to lower the susceptibility perpendicular to the hexagonal axis. The susceptibility parallel to this axis on the other hand remains sensibly constant. The significance of this interesting observation will be discussed in the next section.

(*f*) *Influence of small amounts of zinc*

Zinc and cadmium belong to the same group in the periodic table. Their outermost electronic configurations are similar and both atoms in the neutral state have a spectroscopic state $^1S_0$. The valency electrons for zinc are two in number occupying the 4*s* levels. The atomic radii of zinc and cadmium are 1·53 and 1·73 A respectively.

The zinc used was extra-pure quality from Kahlbaum. Rao (1936*a*), working with the sample used, found the mean susceptibility of the metal from the principal susceptibilities to be 0·167 while direct determinations of polycrystalline rods gave 0·160. The principal susceptibilities for alloys containing small quantities of zinc are given in Table X.

TABLE X

Specimen no.	Percentage by weight of zinc	Relative density	$\chi_\perp$	$\chi_\parallel$	$\chi_\parallel/\chi_\perp$	Mean χ
1	0	8·640	0·163	0·223	1·368	0·183
2	0·29	8·632	0·162	0·222	1·370	0·182
3	0·53	8·624	0·161	0·220	1·366	0·181
4	0·82	8·610	0·162	0·223	1·376	0·182

It will be found on examining Table X that the principal susceptibilities do not alter as a result of introducing small quantities of zinc. Consequently the magnetic anisotropy and the mean susceptibility remain constant.

4. Discussion

The principal diamagnetic susceptibilities obtained for zinc are 0·149 and 0·202 and for cadmium are 0·163 and 0·223. The corresponding atomic susceptibilities are therefore for zinc 9·7 and 13·2 and for cadmium 18·3 and 25·1. The ionic susceptibilities of Zn^{+2} and Cd^{+2} are found by Kido (1933) to be 12·8 and 20·3, while the results of Flordal and Frivold (1935) give values of 13·1 and 21·5 respectively. Applying the rules relating to the calculation of screening constants for different electron shells, Angus (1932) obtained for the ionic susceptibilities of Zn^{+2} and Cd^{+2} the values of 15·45 and 33·94 respectively. Applying the same rules, the contributions of the outermost electrons in the neutral zinc and cadmium atoms to the atomic susceptibility can be calculated. The diamagnetic susceptibility of an electron in any group is given, according to Slater (1930), by

$$= \frac{+0{\cdot}807 \times 10^{-6}(n^*)^2\,(n^* + \tfrac{1}{2})\,(n^* + 1)}{(Z - s)^2},$$

where n^* is the effective quantum number, Z the nuclear charge and s the screening constant for the electron group under consideration.

Calculations based on this relation show that for neutral zinc and cadmium atoms situated free from mutual influences, the diamagnetic susceptibilities per gram atom should be 38·51 and 64·66 since the spectroscopic state of the free atoms is $^1S_0$. Our experimental values are in the neighbourhood of those obtained for the corresponding ions. It can thus be definitely established that the valency electrons contribute very little to the diamagnetism of zinc or cadmium crystals. In fact we have evidence to show that normal to the hexagonal axis, the binding is metallic. The paramagnetic contribution of these electrons works out to 3·2 for cadmium and 3·4 for zinc. It follows from equation (2) in § 1 that the width of the energy band for free electrons has a value of nearly 21 V. This suggests that in zinc and cadmium the electron energy bands are much wider than for free electrons.

These conclusions find support from the specific resistance data for zinc and cadmium crystals along and perpendicular to the hexagonal axis. For comparison the data of Meissner (1926) are shown in Table XI.

This suggests that the conductivity is greater perpendicular to the c axis than along it. The electrons are free to move more easily in the hexagonal plane than normal to it. Thus the contribution to the paramagnetic susceptibility of these electrons is greater normal to the c axis than along it. Thus the magnetic measurements are consistent with electrical conductivity data.

That in zinc and cadmium the electron energy bands are wider than for free electrons receives support from the mass susceptibility of these metals calculated from Landau's formula (equation 3). Table XII is drawn after Mott and Jones (1936).

TABLE XI

Element	Direction with reference to c axis	Specific resistance at 20° C. (10^{-6} units)
Zn	Parallel	5·83
Zn	Perpendicular	5·39
Cd	Parallel	6·8
Cd	Perpendicular	6·5

TABLE XII

Metal	χ_s	χ_l	n_0	χ_L	χ_{n_0}	$\chi_{n_0} + 2\chi_L$
Zn	0·157	0·09	2	0·053	0·200	0·094
Cd	0·18	0·16	2	0·039	0·191	0·113

χ_s = mass susceptibility of solid.
χ_l = mass susceptibility of liquid.
n_0 = number of valency electrons per atom.
χ_L = mass susceptibility according to Landau's formula.
χ_{n_0} = mass susceptibility of ion with n_0 positive charges after Flordal and Frivold (1935).

While the susceptibility calculated in the above manner gives the observed value for zinc in the liquid phase, there is some disagreement for cadmium. It is likely that perfect freedom for the electrons does not obtain in the molten metal.

The effect of stress on a metal in which the binding is metallic has been examined by Honda and Shimizu (1935) who find that in these cases cold-working causes a small increase in the diamagnetic susceptibility. But considering the fact that in zinc, the valency electrons are not perfectly free, such an effect even if it exists must be negligibly small. A similar conclusion is true of cadmium. In the present investigation, no variation of the principal susceptibilities was observed when single crystals of cadmium were subjected to cold-working. Our results are in conformity with those obtained by Rao (1936*b*) who noticed a very small decrease in the case of zinc.

It was mentioned that small quantities of lead lower the susceptibility perpendicular to the hexagonal axis without sensibly affecting the susceptibility parallel to this axis. On the other hand zinc in small quantities does not alter the principal susceptibilities of cadmium. Goetz and Focke (1934) found that the magnetic anisotropy increased when tin was dissolved in bismuth while tellurium decreased this value. Mott and Jones (1936) have drawn attention to this result and accounted for it on considerations of the Brillouin zones constructed for bismuth structure.

Goetz and Focke (1934) also discovered that small quantities of antimony do not alter the magnetic anisotropy. We have an instance of complete atomic solubility in this case. A similar observation is made when small quantities of zinc are dissolved in cadmium. The principal susceptibility and the magnetic anisotropy do not alter as a result of introducing zinc in small amounts. It is likely that the atoms of antimony replace those of bismuth in the hexagonal lattice and at very low concentrations of antimony, the additivity law gives very small changes in the principal diamagnetic susceptibilities. In the same manner when small amounts of zinc are added to cadmium, the zinc atoms might displace some of the cadmium in the crystal. This would not affect appreciably the magnetic constants of the cadmium crystal.

Lead on the other hand has four valency electrons and its presence even in small quantities is bound to disturb the crystalline fields within the crystal. It is found that the susceptibility parallel to the hexagonal axis remains constant while the value perpendicular to this axis decreases. Such a result would occur if the additional valency electrons introduced by the presence of the lead atoms give rise to a larger paramagnetic susceptibility component, in a direction normal to the c axis. Electrical conductivity data with single crystals of cadmium containing small quantities of zinc will be of interest in deciding these problems.

5. SUMMARY

Single crystals of cadmium were prepared by the method of slow cooling and the principal magnetic susceptibilities were determined by the Gouy method. The influence of ferromagnetic impurities on the observed values was eliminated by determining the susceptibilities at different field strengths in accordance with Honda's method. The mean values perpendicular and parallel to the hexagonal axis of the crystal were found to be $-0\cdot163$ and $-0\cdot223$ respectively. The magnetic anisotropy was $1\cdot368$ and the mean diamagnetic susceptibility for a polycrystalline aggregate $-0\cdot183$. The value $-0\cdot163$ of $\chi_\perp$ is close to that of McLennan, Ruedy and Cohen but our value of $\chi_\parallel$ is lower than theirs. This makes the magnetic anisotropy of cadmium much nearer that of zinc than the earlier values indicated.

The influence of cold-working on polycrystalline rods and single crystals of cadmium was investigated. In the case of polycrystalline specimens, a small decrease was obtained. On the other hand, the principal susceptibilities measured with single crystals showed no variation on cold-working.

The influence of small quantities of foreign matter was also a subject of investigation. Lead when added in small amounts affected the principal susceptibility normal to the *c* axis but not the value parallel to this axis. The principal susceptibilities did not alter as a result of adding small quantities of zinc.

These results are examined in the light of the theory of paramagnetism of free electrons developed by Pauli, Landau and Peierls. The valency electrons appear to contribute a paramagnetic component perpendicular to the *c* axis. There is evidence for this from electrical conductivity data. The results obtained with the alloys are examined in the light of these conclusions.

REFERENCES

Angus 1932 *Proc. Roy. Soc.* A, **136**, 569–78.
Bridgman 1925 *Proc. Amer. Acad. Arts Sci.* **60**, 305–83.
Clark 1932 "Applied X-rays", p. 234.
Flordal and Frivold 1935 *Ann. Phys., Lpz.*, **23**, 425–30.
Goetz and Focke 1934 *Phys. Rev.* **45**, 170–99.
Hoge 1935 *Phys. Rev.* **48**, 615–19.
Honda 1910 *Ann. Phys., Lpz.*, **32**, 1027–63.
Honda and Shimizu 1935 *Nature, Lond.*, **136**, 393.
Jackson 1926 *Philos. Trans.* A, **226**, 126–33.
Kido 1933 *Sci. Rep. Tôhoku Univ.* **22**, 385–93.
Landau 1930 *Z. Phys.* **64**, 629–37.
McLennan and Cohen 1929 *Trans. Roy. Soc. Can.* **23**, 159–68.
McLennan, Ruedy and Cohen 1928 *Proc. Roy. Soc.* A, **121**, 9–21.
Meara 1932 *Physics*, **2**, 33–41.
Meissner 1926 *Z. Phys.* **38**, 647–58.
Mott and Jones 1936 "Properties of metals and alloys", p. 211.
Owen 1912 *Ann. Phys., Lpz.*, **37**, 657–99.
Pauli 1926 *Z. Phys.* **41**, 81–102.
Peierls 1933 *Z. Phys.* **80**, 763–91.
Rao 1936*a* *Proc. Ind. Acad. Sci.* **4**, 37–53.
— 1936*b* *Proc. Ind. Acad. Sci.* **4**, 186–203.
Rao and Subramaniam 1936 *Phil. Mag.* **21**, 609–24.
Slater 1930 *Phys. Rev.* **36**, 58–64.
Sone 1920 *Phil. Mag.* **39**, 305–50.

The photosensitivity of diphenylamine-*p*-diazonium sulphate by the method of photometric curves

BY C. F. GOODEVE AND L. J. WOOD

The Sir William Ramsay and Ralph Forster Laboratories, University College, London

(*Communicated by C. K. Ingold, F.R.S.—Received* 12 *February* 1938)

1. "PHOTOSENSITIVITY"

The rate of a photochemical reaction is generally defined as the decrease in the number of molecules of a particular species with time and can be expressed by

$$-\frac{dn}{dt} = \gamma I_A, \qquad (1)$$

where γ is the quantum efficiency and I_A the intensity of the absorbed light in quanta per second. For thin layers of absorbing medium, i.e. layers in which the light intensity is not appreciably reduced, the value of I_A is given by Beer's and Lambert's laws,

$$I_A = \alpha c l I, \qquad (2)$$

where α is the absolute extinction coefficient in cm.^{2} per molecule, c the concentration in molecules per cm.^{3}, l the thickness of the layer in cm., and I the intensity of the incident light. Substituting in equation (1) one obtains an expression for the fraction of the total number of molecules which decompose in unit time,

$$-\frac{dn}{n.dt} = \alpha\gamma\frac{I}{A}, \qquad (3)$$

where A is the area exposed to the light and I/A the light intensity per cm.^{2}.

Equation (3) is applicable to all types of photochemical change taking place under the above conditions. For a particular light intensity per unit area the rate of photochemical change is dependent upon the value of the product $\alpha\gamma$. This quantity we propose to call the "photosensitivity", and, as will be seen from the following examples, it is of greater practical importance than either the extinction coefficient or the quantum efficiency taken singly.*

One of the best known examples of a system very sensitive to light is the reaction between hydrogen and chlorine. Chlorine has a comparatively low extinction coefficient for visible light, $\alpha \doteqdot 10^{-21}$ cm.^{2}. This low extinction is, however, counterbalanced by the fact that the quantum efficiency is generally very high, between 10^3 and 10^5, leading to values of the photosensitivity of 10^{-18} to 10^{-16} cm.^{2} per quantum absorbed. Solutions of visual purple (see § 3) are very photosensitive. Dartnall, Goodeve and Lythgoe (1936, 1938; hereinafter referred to as "previous papers") have shown that this arises from the fact that the extinction coefficient is high, while the quantum efficiency has a value of about unity. The photosensitivity of this substance is about 10^{-16}. On the other hand, there are reactions such as the photodecomposition of aliphatic nitroso compounds (Hammick and others 1935, 1937) in which the quantum efficiency is unity, but the extinction coefficient very low, the value of the photosensitivity being less than 10^{-19}. If a solution of a nitroso compound is exposed side by side with a solution of visual purple, it decomposes at a rate 1000 times slower than the visual purple, although its quantum efficiency is about the same. In a further class there exist many dyestuffs of a very high extinction but of a low photosensitivity arising from the fact that their quantum efficiencies are extremely small.

Photochemical reactions behave differently with different wave-lengths of incident light. As quantum efficiencies are often more or less independent of wave-length it is an advantage to have a term "photosensitivity" defined in such a way as to be a function of wave-length. For example, visual purple has a high sensitivity to green light, while the diazonium salt discussed below is insensitive to this light and has a high sensitivity in the near ultraviolet. Their quantum efficiencies are about the same but their photosensitivities follow their absorption curves; visual purple absorbs green light strongly, while the diazonium salt is transparent to light of this wave-length.

---

* Many equations for the kinetics of photochemical reactions have been proposed and may be found in text-books on photochemistry. Most of these, however, are based upon the use of *energy* units for the light intensity. Henri and Wurmser (1913), for example, called the constant in their reaction velocity equation the "susceptibilité photochimique", and showed the importance of the extinction coefficient in determining its value.

2. THE METHOD OF PHOTOMETRIC CURVES

Equation (3) has been derived for thin layers of absorbing substance. In the previous paper (1936) it was shown that if Beer's and Lambert's laws are obeyed, and if the quantum efficiency is independent of the concentration, the following equation describes the increase with time of the intensity of the transmitted light, I_t,

$$\log_e \frac{I_t}{I - I_t} = \frac{\alpha\gamma I}{A}\, t + \text{constant}. \tag{4}$$

The terms I and I_t on the left-hand side of the equation may be expressed in any convenient units such as galvanometer deflexions, where these are linear with light intensity, but I on the right-hand side must be in quanta per second. If a photometer is used giving readings of density, the values of the left-hand side of the equation can be calculated from the relation

$$\log \frac{I_t}{I - I_t} = -\log(\text{antilog density} - 1). \tag{5}$$

It was further shown (1936, 1938) that if either coloured photostable impurities are present or coloured stable products are formed, the equation becomes

$$\log_e \frac{I_t}{I_f - I_t} = \phi \frac{\alpha\gamma I}{A}\, t + \text{constant}, \tag{6}$$

where
$$\phi = \frac{I_f}{I_f - I_t} \, \frac{I - I_t}{I} \, \frac{\log I_f/I_t}{\log I/I_t}, \tag{7}$$

and I_f is the final value of I_t.

In the second paper (1938) the value of ϕ was shown to be practically constant for a particular experiment, its value depending upon the ratio I_f/I. If, therefore, the logarithmic function of the light intensities, as given on the left-hand side of equation (6), is plotted against the time, a straight line should be obtained of slope $\phi \frac{\alpha\gamma I}{A}$. Measurements of I and A and of the optical density or of the transmitted light intensity during the course of a photochemical change can thus be used to determine the photosensitivity.

Equation (6) is a completely general equation expressing the progress of any simple photochemical change. This equation and the above graphical method play the same part in photochemistry as do the equations and the corresponding straight-line graphs for first- and second-order reactions in reaction kinetics (these latter equations are $\log C = -k_1 t + \text{constant}$ and

$1/C = k_2 t + \text{constant}$, respectively). The "photosensitivity" then corresponds to the "reaction velocity constants" k_1 and k_2.

The "method of photometric curves" permits an examination of a photochemical reaction throughout its whole course. It is very sensitive to the presence of consecutive reactions and other complications.

3. APPLICATIONS

The above equations have been applied extensively in the previous papers (1936, 1938) to the photodecomposition of visual purple, a light-sensitive coloured substance extracted from retinae of animals which were "dark-adapted" before death. This application raises an interesting point. The dimensions of photosensitivity, $\alpha\gamma$, are cm.^{2} (converted per quantum absorbed), i.e. they do not contain "concentration". Equation (6) can therefore be applied to a solution the concentration of which is unknown and of course gives by itself no information from which the concentration could be determined. At the present time there is no *chemical* way by means of which the molecule of visual purple may be defined or its amount or concentration determined. The value of $\alpha\gamma$ may be measured but not that of either α or γ separately.

From arguments based upon simple photochemical generalizations given in the previous papers it was deduced that the value of γ was equal to or not much less than unity. On this basis a probable absolute scale for the extinction coefficient curve was obtained, thus making it possible to calculate the concentration of any particular solution.

It was, however, highly desirable to test these photochemical generalizations on a known system bearing the closest possible resemblance to visual purple. Many dyestuffs and coloured organic compounds were tested, but all were found to have a very much lower sensitivity or were thermally unstable. Of a number of substances, suggested and kindly supplied by Mr D. W. Powell, diphenylamine-*p*-diazonium sulphate was found to have the required properties, although its sensitivity was confined to the near ultra-violet.

4. EXPERIMENTAL

The design of the apparatus was basically the same as described previously (1936, 1938). A quartz monochromator was used with an atmospheric pressure mercury arc almost touching the entry slit. Inside this slit was fixed a piece of Wood's ultra-violet glass to filter off most of the radiation other than that of 365 mμ (see § 6). The radiation leaving the exit slit and falling on the cell was found spectroscopically to be at least 99 % pure, there being small intensities of wave-lengths 313 and 334 mμ. An additional lens was fixed on the side of the cell holder near to the monochromator in order to make the light passing through the cell as nearly parallel as possible, and thus to reduce refraction of the light on to the side of the cell. The light transmitted by the solution was allowed to fall on a Weston photocell, which was connected to a Zernicke galvanometer the deflexions of which were found to be proportional to the light intensity. The bleaching was interrupted from time to time, and the incident light intensity, I_m, was measured. Absolute calibrations were made by means of a thermopile in the usual manner.

Although the mercury lamp was run from a constant voltage supply its intensity varied slightly. The variations were found to be proportional to the current through the lamp over a narrow range and were therefore corrected from measurements of this current taken simultaneously with the galvanometer deflexions.

The sensitivity of the photocell was found to decrease (as compared with the thermopile) during the course of a run. The fatigue was regular, and corrections were made for this by periodic calibrations.

The intensities of the light given in equation (6) refer to those incident upon or transmitted by the solution. It is not possible to measure these light intensities, but, by making allowance for the reflected light, they can be calculated from the light incident upon or transmitted by the cell. No corrections need be made to the intensity terms on the left-hand side of equation (6), since the same correction would apply in the numerator as in the denominator. The light intensity which appears on the right of the equation must, however, be replaced by the following expression:

$$I = \beta I_m(1+\psi), \tag{8}$$

where β is the fraction of light transmitted by a quartz plate, one side of which is in contact with air and the other in contact with water or dilute aqueous solution. ψ is a term to account for the light which, reflected back into the solution from the second quartz face, causes additional photochemical change. At the beginning of an experiment with a concentrated solution it will be very small, but it increases as the bleaching proceeds. By simple optics it can be shown that

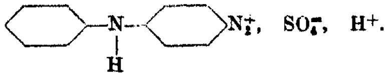
$$\psi = \frac{1-\beta}{\beta^2}\frac{I'_t}{I_m}, \tag{9}$$

where I'_t is the light transmitted from the second quartz face. The value of ψ increases with I'_t, but fortunately ϕ decreases and the product $\phi(1+\psi)$ remains constant to about 1 % during the essential part of the run. A mean value of this product was used in the combination of equations (6) and (8).

5. THE DIAZONIUM SALT

The photochemical instability of diazonium salts has often been described in the literature (Murray 1933; Eder 1934; Henriquez 1936). The decomposition of the salt, diphenylamine-*p*-diazonium sulphate, used in the present work was studied by Ikuta (1888) and Schröter (1930). The former found on analysis that it had the formula $C_6H_5.NH.C_6H_4.N_2.SO_4H$. The sample used for the work described below was obtained in a high state of purity and was further purified by recrystallization from a 3 : 1 alcohol-water mixture. It was subjected to a chemical analysis with the following results: carbon 49·3 % (49·15), nitrogen 14·0 % (14·3), hydrogen 3·8 % (3·75) (the figures in parentheses give the percentages calculated from the above formula).

Diazonium salts are generally completely ionized in solutions of the strength used here, i.e. less than 10^{-4} N (Hantzsch 1895). It would appear, therefore, that the solution contains the following species:



Determinations of the *p*H of a number of solutions by means of a Cambridge Electrometer Valve *p*H Meter (glass electrode) confirmed the ionization of one equivalent of hydrogen.

6. THE ABSORPTION SPECTRA

The absorption spectra, both of a pure solution of the diazonium salt and of the same solution after bleaching, were measured on a Hilger "Spekker" Photometer combined with a Hilger E 3 Spectrograph. The results are shown in fig. 1. The molar extinction coefficient, ϵ, is defined by the equation

$$\log_{10} I/I_t = \epsilon Cl, \tag{10}$$

where C is the concentration in g. mol./l.

It is seen that the absorption of the diazonium salt possesses a marked maximum at 26,300 cm.$^{-1}$ (380 mμ), and that the product of bleaching is comparatively transparent at frequencies less than 30,000 cm.$^{-1}$ (330 mμ). It was therefore found convenient to bleach with light of 365 mμ from a mercury arc lamp.

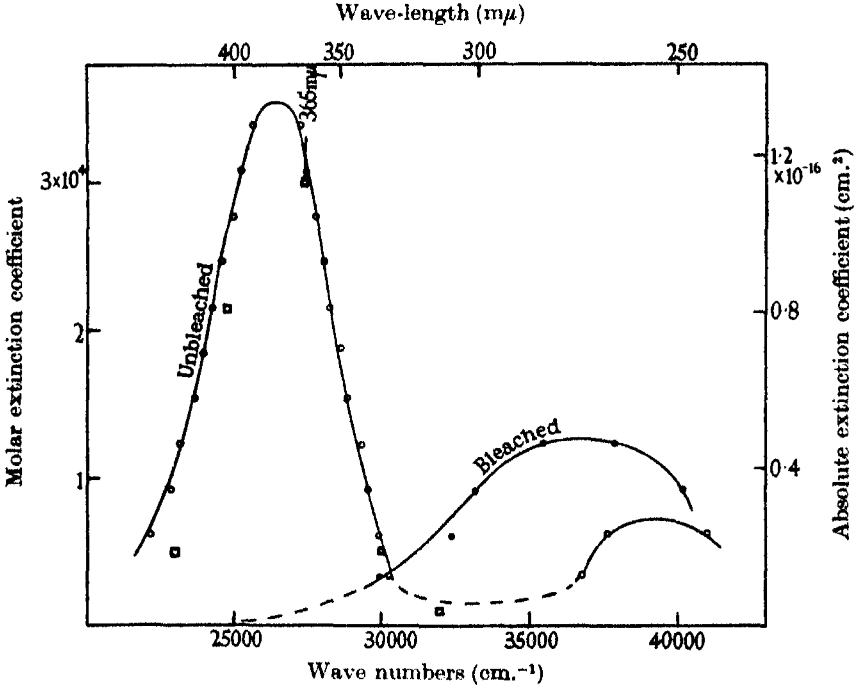


Fig. 1. The absorption curves of unbleached and bleached diphenylamine-*p*-diazonium sulphate. ☐ observations by Schröter.

7. Results

The experimental values of $\log_{10}\dfrac{I_t}{I_f - I_t}$ for sixteen experiments are plotted against the time in fig. 2. The area of the cell exposed to the light, A, was 0·785 cm.^{2}. Other data in connexion with these experiments are given in Table I. All but four of the experiments are recorded, these four being rejected owing to experimental mishap. A typical galvanometer deflexion curve, that for Exp. 4, is shown by full dots.

It is seen that the experimental results when plotted as the logarithmic function lie accurately upon straight lines except in the upper part of some of the curves where the error in the logarithmic function is grossly disproportionate to the error in the deflexions themselves. The fact that these lines are straight indicates that the photosensitivity is constant throughout the course of the bleaching and thus independent of concentration. The above method of analysis is therefore applicable. The quantities required for the calculation of $\alpha\gamma$ are given in Table I in columns 3–6. The values of the extinction coefficient, α, given in column 9, are calculated from the initial densities of the solutions. Using the mean of these the values of the quantum efficiency have been calculated and are given in column 10.

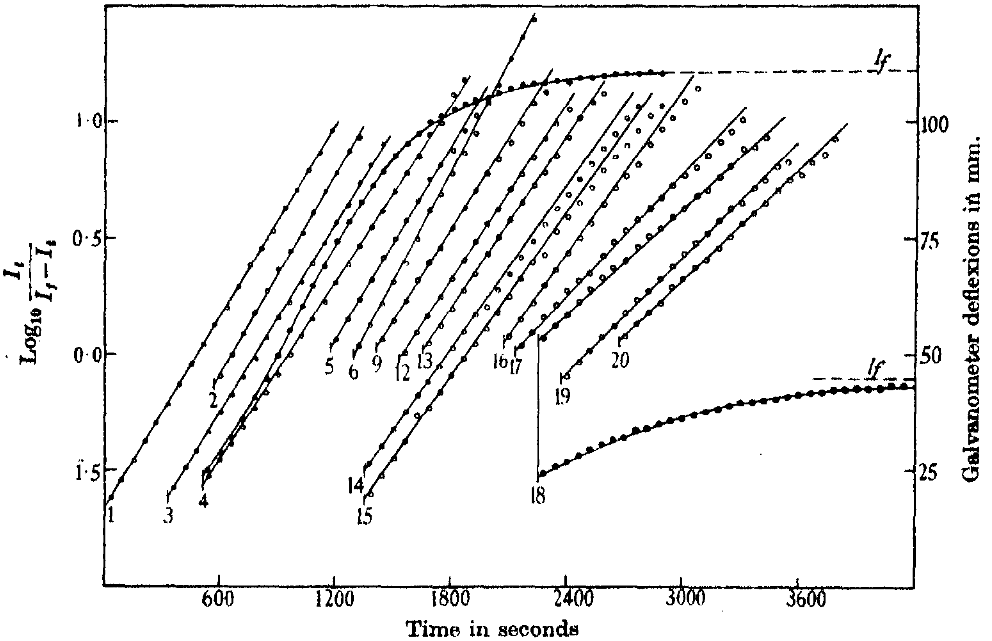


Fig. 2

The first ten experiments were carried out over a wide range of temperature, and it may be seen from the results that the photosensitivity, the extinction coefficient, and the quantum efficiency are all independent of temperature. The importance of this fact from a general photochemical point of view is discussed below.

Exps. 15 and 16 were made to determine whether or not the bleaching was dependent on the presence of air dissolved in the solution. The cell used for these two experiments was fitted with a bulb in the stem. After filling it was connected to a rotary Hyvac pump and evacuated until about one-third of the solvent had boiled away. The bleaching was then carried out under normal conditions. The fact that the photosensitivity is unaffected by the removal of air indicates that the bleaching process is not an oxidation involving oxygen.

In order to test the accuracy of the correction factor ϕ, a photostable yellow tartrazine dye—1110 Baeyer—was dissolved in two solutions of the diazonium salt to act as an internal filter. One solution was used for Exps. 17 and 19, and the other for 18 and 20. The concentrations of the tartrazine dye are given in brackets in column 2. The dye absorbed about the same amount of light as did the initial amount of diazonium salt. The deflexion-time curve for Exp. 18 is also shown by full dots in fig. 2. The final deflexion as shown by the broken line is very much less than that for the pure solutions owing to the presence of the tartrazine dye. This low value of the final light intensity, I_f, produces a low value of ϕ, which compensates for the decreased slope of the straight lines obtained from these experiments. It appears from Table I that the photosensitivity is unaltered by the presence of the dye, a fact which indicates that the dye does not sensitize the reaction and that the ϕ correction provides full compensation unless by chance the photosensitivity is increased by a factor inversely proportional to an error in ϕ. This latter possibility is most unlikely.

TABLE I

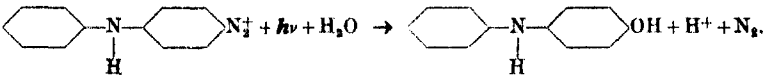
1 Exp. no.	2 Initial conc. $\times 10^5$, N	3 Slope $\times 2{\cdot}303$	4 Intensity on cell, quanta/sec. $\times 10^{-13}$	5 ϕ	6 $1+\psi$	7 Temp. °C.	8 $\alpha\gamma \times 10^{17}$	9 $\alpha \times 10^{16}$	10 γ
1	5·45	0·00318	7·179	0·978	1·025	20·0	3·68	1·07	0·338
2	2·725	0·00337	7·035	1·000	1·035	20·0	3·85	1·04	0·353
3	5·45	0·00312	6·983	0·967	1·025	45·4	3·75	1·03	0·344
4	5·37	0·00292	7·007	0·944	1·025	49·9	3·58	1·03	0·328
5	2·10	0·00321	6·784	0·986	1·038	15·0	3·84	1·09	0·352
6	2·18	0·00366	7·806	0·976	1·039	30·0	3·85	1·13	0·353
9	2·16	0·00306	7·466	0·982	1·038	30·05	3·34	1·07	0·306
12	2·19	0·00290	6·549	0·984	1·039	20·8	3·60	1·11	0·330
13	2·19	0·00288	6·182	0·944	1·035	44·9	3·97	1·20	0·364
14	4·38	0·00272	6·170	0·970	1·021	22·8	3·70	1·13	0·339
15	—*	0·00272	6·372	0·977	1·025	20·2	3·54	—	0·325
16	—*	0·00272	6·492	0·965	1·033	20·3	3·49	—	0·320
17	1·93 (9·3)†	0·00205	6·201	0·751	1·019	20·15	3·59	—	0·329
18	1·84 (13·1)†	0·00174	6·163	0·684	1·017	20·15	3·37	—	0·309
19	1·93 (9·3)†	0·00194	6·069	0·750	1·018	20·15	3·48	—	0·319
20	1·84 (13·1)†	0·00184	5·880	0·686	1·017	20·1	3·73	—	0·342
		Mean of experiments nos. 1–16					3·68	1·09	0·34 ± 0·02

* Solution evacuated before experiment to remove air.
† Concentration of tartrazine dye solution added.

All of the above experiments were carried out with unbuffered solutions and at an initial pH between 4·2 and 4·7. Rough experiments were carried out in solutions buffered to pH's 6 and 10. The photosensitivity was found to be about the same at *p*H 6 as at 4·2, but about three times higher at *p*H 10. The extinction coefficient for $\lambda = 365$ mμ was, however, independent of *p*H. It would appear, therefore, that the quantum efficiency of 0·34 for acid solutions rises to unity for alkaline solutions.

8. DISCUSSION

Schröter (1930) carried out a number of experiments on the bleaching of diphenylamine-*p*-diazonium sulphate, using a simple form of apparatus, and found the quantum efficiency to be more or less independent of wave-length and concentration over the narrow ranges studied. His values for the quantum efficiency were 0·6 from measurements of the change of optical density and 0·2 from measurements of the nitrogen evolved, and he suggested as an explanation of the discrepancy that the reaction is complicated. On the other hand, both Schmidt and Maier (1931) and Horio and Yamashita (1935) found on irradiation of this salt that the nitrogen evolved was equal in quantity to the diazonium nitrogen present. The photochemistry of a number of diazonium salts has been studied by Orton and Coates (1907), who found that phenols were produced equivalent to the original amounts of salt. We may therefore represent our reaction as



This mechanism has been confirmed in the present work by measurements of the hydrogen-ion concentration before and after bleaching. Full precautions necessary for accurate determinations of unbuffered solutions were not taken, but in general it was found that the hydrogen-ion concentration increased by a factor of about 2 during the bleaching process. It would appear, therefore, that the discrepancy in Schröter's experiments was due to experimental error.

It is of interest to discuss the above reaction together with the bleaching of visual purple in the light of some general photochemical principles postulated by Franck and Rabinowitch (1934) and shown by Norrish (1937) to be of wide application. According to these authors, reactions in solution which are independent of temperature and concentration do not involve atoms or free radicals except when these react quickly with the solvent. In other words, the reaction goes to completion in a period of time of the order of magnitude of the interval between collisions by the solvent molecules with the activated reactant. In such cases it is impossible for chain reactions to occur and therefore impossible for the quantum efficiency to be greater than unity. The photodecomposition of the diazonium salt conforms with this generalization. The fact that the bleaching of visual purple is independent of temperature and concentration (see previous papers, 1936, 1938) leads to the conclusion that the quantum efficiency for this process is also not greater than unity.

The photochemical reactions of the diazonium salt and visual purple have several points in common, although of course it is not suggested that there is any chemical relation between them. Their absorption curves are continuous and of an almost identical shape and their photosensitivities are of the same order of magnitude. The maximum extinction coefficient of the diazonium salt is normal for continuous absorption curves. If visual purple also has a normal extinction coefficient, its quantum efficiency will be of the order of magnitude of unity.

The authors are indebted to Dr R. J. Lythgoe for his advice and criticism, to Dr E. E. Schneider and Mr J. A. Kitchener for assistance with the experiments and to the Rockefeller Foundation for a generous grant in aid of this research.

SUMMARY

The method of photometric curves, used previously to measure the photosensitivity of visual purple solutions, has been applied to the bleaching of diphenylamine-*p*-diazonium sulphate.

The diazonium salt was bleached with light of wave-length 365 mμ, and it was found to have a quantum efficiency of $0{\cdot}34 \pm 0{\cdot}02$, independent of concentration and temperature. The photosensitivity was also found to be unaffected by the removal of dissolved oxygen and by the addition of an internal filter.

These results have been compared with those obtained with visual purple, and their photochemical significance discussed.

REFERENCES

Dartnall, H. J. A., Goodeve, C. F. and Lythgoe, R. J. 1936 *Proc. Roy. Soc.* A, **156**, 158–70.

—— —— —— 1938 *Proc. Roy. Soc.* A, **164**, 216–30.

Eder, J. M. 1934 *Z. wiss. Photogr.* **33**, 1–12.

Franck, J. and Rabinowitch, E. 1934 *Trans. Faraday Soc.* **30**, 120–31.

Hammick, D. Ll., Anderson, K. D. and Crumpler, C. J. 1935 *J. Chem. Soc.* pp. 1679–84.
Hammick, D. Ll. and Lister, M. W. 1937 *J. Chem. Soc.* pp. 489–93.
Hantzsch, A. 1895 *Ber. dtsch. chem. Ges.* **28**, 1734–53.
Henri, V. and Wurmser, R. 1913 *J. Phys. théor. appl.* **3**, 305–23.
Henriquez, P. C. 1936 *Chem. Weekbl.* **33**, 242–6.
Horio, M. and Yamashita, S. 1935 *Z. wiss. Photogr.* **33**, 273–80.
Ikuta, M. 1888 *Ann. Chem.* **243**, 272–89.
Murray, H. D. 1933 *Photogr. J.* **57**. April supplement, pp. 6–16.
Norrish, R. G. W. 1937 *Trans. Faraday Soc.* **33**, 1521–8.
Orton, K. J. P. and Coates, J. E. 1907 *Trans. Chem. Soc.* **91**, 35–56.
Schmidt, J. and Maier, W. 1931 *Ber. dtsch. chem. Ges.* **64**, 778–9.
Schröter, W. 1930 *Z. wiss. Photogr.* **28**, 1–29.

An X-ray study of the iron-nickel-aluminium ternary equilibrium diagram

By A. J. Bradley, D.Sc., *Royal Society Warren Research Fellow*,
and A. Taylor, Ph.D., *University of Manchester*

(*Communicated by W. L. Bragg, F.R.S.—Received* 18 *February* 1938)

Stimulated by the special interest of iron-nickel-aluminium alloys for the permanent magnet industry, we have made an X-ray investigation of this system. The information obtained has led to a new view of the nature of the permanent magnets of iron, nickel and aluminium, which has been briefly described elsewhere (Bradley and Taylor 1937 *a*, *b*, *c*). In the present paper we confine ourselves to a description of the X-ray results, which have enabled us to draw a complete ternary diagram representing the phases formed by these metals.

The iron-nickel-aluminium alloys are closely related to the systems FeAl, NiAl, CoAl. Every phase which occurs in the ternary system has an analogy in one or another of these binary systems. This makes it possible to interpret the present results in the light of our previous knowledge of those systems.

Up to 50 atomic % of aluminium, there are only face-centred cubic and body-centred cubic structures. Interesting features of these alloys are the formation of superlattices and the phase equilibria. The face-centred cubic phase field is not merely separated from the body-centred cubic phase field by a two-phase area, as one might suppose. The peculiar shape of the β (body-centred cubic) phase gives rise to a two-phase field in which alloys have two types of body-centred cubic lattice with widely different compositions. There are, in all, three two-phase areas and a three-phase area lying between the face-centred cubic and body-centred cubic phase fields. In addition the face-centred cubic region itself encloses a miscibility gap.

The body-centred cubic alloys are of three types. The iron-rich alloys have no superlattice; near Fe_3Al the superlattice is of the type X_3Y, and with more aluminium the alloys have superlattices like FeAl or NiAl. Even where the superlattices are of the same type, there are remarkable variations in lattice spacing. These irregularities make it difficult to fix the phase boundaries in the usual way, but the strength of the superlattice lines acts as a guide.

Beyond 50 atomic % of aluminium the system becomes more complex. Whereas the aluminium-poor alloys contain only face-centred cubic and body-centred cubic structures, the aluminium-rich alloys include eight other phases, some of them having very complicated structures. Four of these phases occur in the iron-aluminium system; two occur in the nickel-aluminium system; while the remaining two are related to structures found in cobalt-aluminium alloys. A more detailed account of the aluminium-rich alloys will be given in a later paper. At present we shall deal with the nickel-iron end of the system, up to the boundaries of the body-centred cubic phase field.

Before giving the results of the present investigation, it will be advisable to review recent work on the three binary systems, Fe-Ni, Ni-Al, Fe-Al.

THE IRON-NICKEL SYSTEM

The iron-nickel system has in the last two years been the subject of three independent researches, in each of which X-ray methods have been used. Jette and Foote (1936) measured the lattice spacings of alloys after quenching from a series of temperatures. Bradley, Jay and Taylor (1937) have measured the lattice spacings both after quenching and after slow cooling. Owen and Yates (1937) have measured the spacings at elevated temperatures. The general result of the X-ray work is to confirm the phase diagram given by Merica (1930), which is given in fig. 1. This shows the existence of two phase fields, one face-centred cubic stretching from pure nickel towards the iron end of the diagram, and the other body-centred cubic, including pure iron and dissolving varying amounts of nickel at different temperatures.

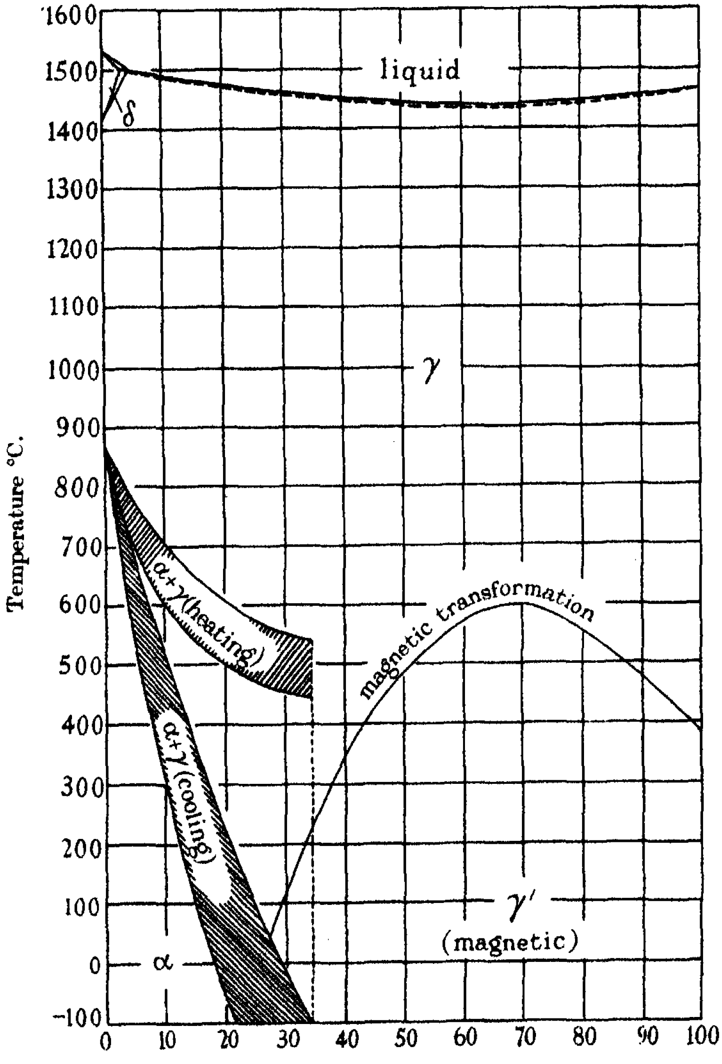


Fig. 1. The iron-nickel system (Merica).

The exact position of the boundaries is still in doubt. This is due to a peculiar hysteresis effect in the iron-rich alloys. The face-centred cubic phase field stretches right across the diagram at temperatures above

870° C. Below this temperature, the iron-rich alloys begin to change to the body-centred cubic form. The progress of the transformation depends on the exact way in which the metal is cooled. Fig. 1 shows that the transformation occurs at a much lower temperature on cooling than on heating.

Alloys containing less than 6 % of nickel change completely to the body-centred cubic form on cooling. With more nickel it soon becomes difficult to complete the transformation. Slow cooling then gives a mixture of body-centred and face-centred cubic patterns. Beyond 29 % of nickel the face-centred cubic pattern is obtained alone. We will here consider the system as containing a two-phase region extending from about 4·4 % of nickel to 29 % of nickel because this is the result of the heat treatment which we have employed, but we make no claim that this is the state of equilibrium, this difficult question being still under discussion. With this simplification, it is possible to reconcile the results obtained by slowly cooling alloys belonging to different ternary systems.

THE NICKEL-ALUMINIUM SYSTEM

The nickel-aluminium system has recently been re-examined independently by X-ray and by thermal and microscopic methods. Our X-ray powder photographs (Bradley and Taylor 1937 *a*) confirm the broad outlines of the phase diagram proposed by Gwyer (1908), but an additional phase (α') was found and the boundaries of the other phases were considerably modified. The crystal structure of each phase was solved and some unexpected features were found. It will be convenient to refer to these when discussing the corresponding portion of the ternary system.

The X-ray results lead to the modification of the nickel-aluminium diagram given by us previously (Bradley and Taylor 1937 *a*, p. 57). This is substantially the same as the diagram published a few months later by Alexander and Vaughan (1937), using thermal and microscopic methods. The only important difference is near the composition Ni_3Al at temperatures above 1000° C. We found a greater solubility of aluminium in the α-phase at 1100° C., which suggested that the two face-centred cubic phase fields (α and α') might join together at high temperatures. The microscopic evidence did not confirm this. However, still a third interpretation of this part of the diagram has been advanced (Nishimura and Watanabe 1936). Whichever view may prove to be correct, there can be no doubt of the accuracy of this diagram at temperatures below 1000° C., where it is substantially in agreement with Alexander and Vaughan's diagram.

For comparison with the present results, we need only mention the structures of the following phases: α face-centred cubic, α′ face-centred cubic with a superlattice (Ni_3Al), β body-centred cubic with a superlattice (NiAl).

THE IRON-ALUMINIUM SYSTEM

The iron-aluminium system has been investigated a number of times, but so far no diagram has been produced capable of explaining all the known facts, including the X-ray work on the iron-rich alloys (Bradley and Jay 1932*a*, *b*) and the recent magnetic measurements (Sykes and Evans 1935). We therefore suggest the diagram given in fig. 2 which is based on Ageew and Vher's diagram (1930) with the modifications which are necessary to bring it into line with later developments.

In earlier diagrams, the letters α and δ have been used for the same phase field. This includes the composition FeAl, which is exactly like NiAl and CoAl in structure. The analogy is obscured by using different letters for the three phases. If all three are called β it is easier to understand the ternary diagrams. For example, in the quaternary system Fe-Cu-Ni-Al, which we are investigating, FeAl, NiAl and Cu_3Al all appear to be linked together as parts of a single body-centred cubic phase field. It is more convenient to call the body-centred cubic structures β whatever the composition rather than to use a variety of terminology for different corners of the system. Similarly, the nickel-copper system links up with α-Cu-Al and α-Ni-Al, and with the other face-centred cubic phase fields. We have found it convenient to call all the face-centred cubic structures α, reserving the use of γ for complex types of structure similar to Cu_9Al_4.

The large areas of face-centred cubic and body-centred cubic structures may be subdivided into a number of portions with definite boundaries. These are not separate phase fields, not being separated by two-phase regions. These boundaries indicate either changes in atomic arrangement (order-disorder changes) or changes in the magnetic state (from ferromagnetic to non-magnetic). It is convenient to indicate each region by means of suffixes. *m* and *n* indicate the magnetic and non-magnetic portions; 1 and 2 indicate ordered structures of the type Fe_3Al and FeAl respectively.

In fig. 2, the boundary of the β-phase is taken from our X-ray measurements; it agrees well with the original work of Gwyer and Phillips (1927). The only other feature of the diagram which calls for comment is the behaviour of $FeAl_3$. We have found that this phase is sensitive to heat treatment. After quenching from 600° C., it was almost a single phase.

Slow cooling converted it into a mixture of Fe_2Al_5 and a phase containing a little more aluminium than $FeAl_3$. An X-ray examination of other alloys on the aluminium-rich side of $FeAl_3$ confirmed this. An alloy of the composition Fe_2Al_7 was a mixture of $FeAl_3$ and aluminium after quenching from 600° C. Slow cooling converted it into what was apparently a single phase. The pattern was almost identical with that of quenched $FeAl_3$; there were however slight differences in the positions of the lines, suggesting changes in lattice spacing and in ordering.

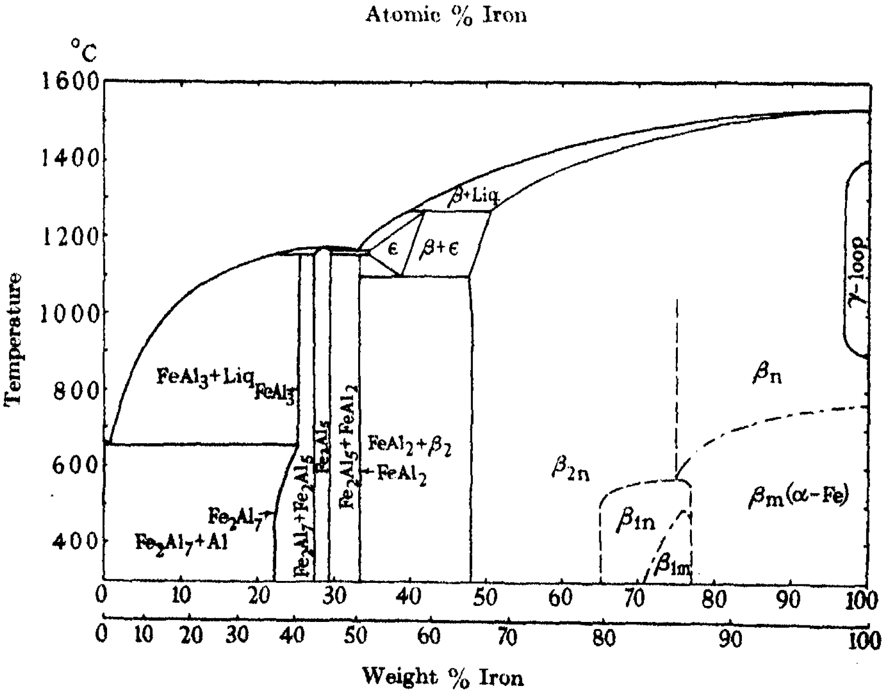


FIG. 2. Iron-aluminium diagram.

It is not quite certain from our present evidence that quenched $FeAl_3$ and slowly cooled Fe_2Al_7 are in the same phase field. If so, $FeAl_3$ represents its composition from the melting point probably down to 650° C., while Fe_2Al_7 is correct for much lower temperatures. Careful experiments, made by the courtesy of Dr Haughton, showed no evidence of a discontinuity in the cooling curve.

THE TERNARY SYSTEM

At the outset of this investigation our knowledge of the ternary Fe-Ni-Al system was entirely derived from the important pioneer investigation of Köster (1934), which explained the relationship of the face-centred cubic and body-centred cubic phase fields. It included a very thorough study of the liquidus by thermal analysis, showing the gradual change from the eutectic transformation in the nickel-aluminium system, to the peritectic transformation in the iron-nickel system. A microscopic study fixed the boundaries in the solid state, showing the existence of a wide area of solid solubility stretching from pure iron to the compound NiAl. The Curie points were also determined.

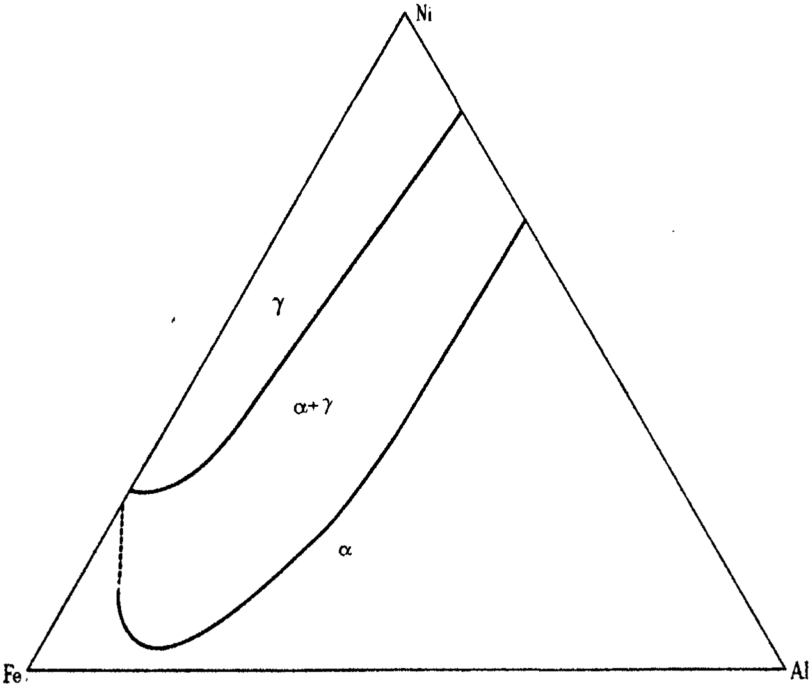


FIG. 3. Iron-nickel-aluminium system (after Köster).

An isothermal phase-diagram for room temperatures, taken from Köster's paper, is given in fig. 3, in which the compositions are represented atomically for comparison with the present results. The boundaries of the face-centred cubic and body-centred cubic phase fields obtained microscopically were found to be extremely sensitive to heat treatment. Eight vertical sections were made through the ternary model, each of which showed that the solid solubility for both phases was much greater at high temperatures, than after slow cooling.

The present investigation is confined to a study of the phase equilibria at lower temperatures, where we have been able to amplify the information given in fig. 3. We confirm the general trend of the boundaries of the two phase fields, but show a much more complicated region between them.*

EXPERIMENTAL METHODS

We have examined a very large number of alloys from all parts of the ternary system. In each case 30 g. of the pure metals was weighed out and melted down in a high-frequency induction furnace under a low pressure of hydrogen, the gas being pumped off before the alloys could solidify. The materials used were Hilger iron and nickel (batch nos. 8038 and 7636), purity 99·97 %, and aluminium from the British Aluminium Company and Aluminium Française of purity 99·992 %. The melt was made in a crucible lined with alumina.

The alloys were lump-annealed and powder-annealed in each case. The lump-annealing was carried out as near to the melting point as possible, to ensure homogeneity. The powder-annealing was done by slow cooling *in vacuo*, the object being to get a representative picture of the conditions below 600° C. The results are probably not very different from what would be obtained by quenching from 500° C., but the technique is far simpler. This is very important, because the results from one batch of alloys were used to decide the compositions to make up in the next batch. It would have been impossible to have given a lengthy heat treatment under these circumstances.

At first only a few alloys were made up, with widely different compositions. From their X-ray powder photographs a rough idea was obtained of the probable positions of the single-phase areas. More alloys were then made up, with the object of defining these areas more exactly. By a process of trial and error, the phase boundaries gradually assumed their final state. It was found that sufficiently good results could be obtained with composition intervals of about 2·5 atomic % even where there were features of special interest. Where the diagram was less detailed less alloys were required. It is instructive to note that if the alloys had all been made up in advance at these intervals, 741 specimens would have been needed. By using the method of trial and error this number was reduced to 130. Most of these specimens are situated near to phase boundaries.

---

* We have just received an account of the structure of nickel-aluminium-magnet-steel, which agrees with the present work (Kiuti 1937).

Each powder photograph checks up the lump annealing. If there is segregation, a bad photograph is obtained. The lines are diffuse, showing that the material does not have a uniform lattice spacing. Sometimes the phase rule is not obeyed. There may be four or more phases present. Such discrepancies are usually removed by changing the temperature of lump annealing. Only in exceptional cases, where there is a reaction occurring below 600° C., does this method not lead to a complete solution.

In the final heat treatment, the alloys were lump-annealed for about three days at 1300° C. unless otherwise stated. The powders were cooled down to room temperature at about 10° C. per hour starting at 900° C. unless another temperature is given in the appropriate table.

The X-ray powder photographs were taken in a camera of 9 cm. diameter, using either Cu $K\alpha$ radiation or Co $K\alpha$ radiation. The choice was based on considerations of wave-length, absorption coefficient and atomic scattering factor. Cobalt gives accurate spacing measurements for all the body-centred cubic alloys, on account of the high angle of the 310 reflexion. The absorption coefficient is low, especially for the iron-rich alloys. A specially useful feature is the ability to distinguish between iron and nickel atoms, in spite of the small difference in atomic number. For most other radiations the scattering power of nickel is only about 1 part in 14 greater than that of iron, with cobalt radiation the difference may amount to as much as 25 % at high angles.

Copper radiation has a more limited application. The shorter wave-length is disadvantageous for the more complicated structures, which are found in the aluminium-rich alloys. Its chief use was for the lattice spacing measurements of the face-centred cubic alloys, which do not give suitable lines with cobalt radiation.

THE TERNARY DIAGRAM

The ternary diagram for the whole Fe-Ni-Al system is given in fig. 4. The circles represent the compositions of alloys which have been investigated, the scale being in atomic percentages. Different areas are shown in the following way: single-phase, two-phase, and three-phase fields are separated by definite boundaries. Single-phase fields have curved boundaries and are left blank; in two-phase fields the directions of the tie-lines are indicated; three-phase fields are unshaded triangles. Within some of the single-phase fields, there are subdivisions which refer to the superlattices and ferromagnetic properties. The limits of superlattice structures are shown by broken lines; the magnetic boundaries, which are the limits of ferromagnetic properties at room temperatures, are indicated by chain-dotted lines. The diagram includes a key to the nomenclature of the phases, based on the type of structure.

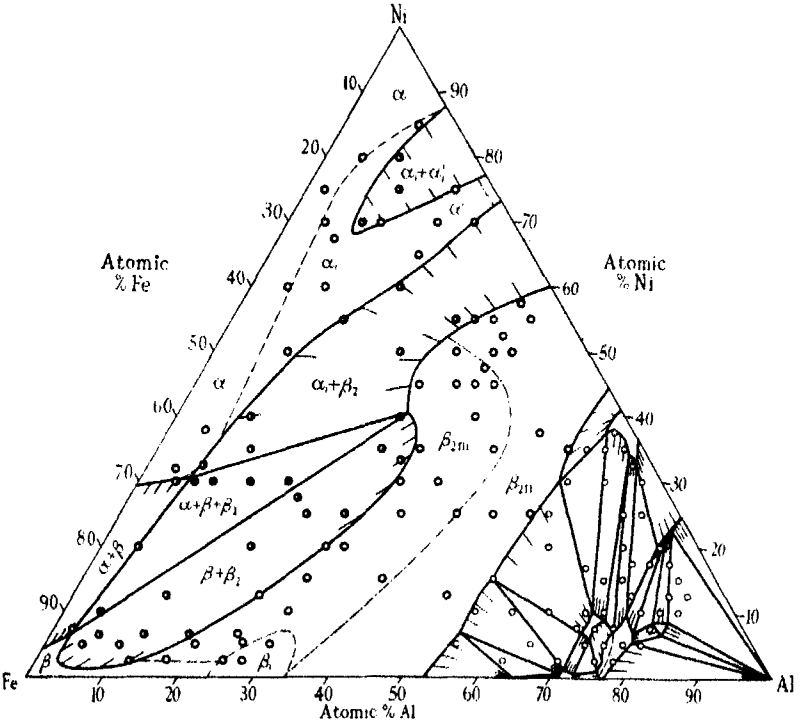


FIG. 4. The iron-nickel-aluminium system. Alloys cooled at 10° C. per hour.

— — — — superlattice boundaries;
· — · — · magnetic boundaries;

○ single-phase alloys;
◑ two-phase alloys;
● three-phase alloys;

α face-centred cube;
α_1 superlattice like Ni_3Al;
β body centred cube;
β_1 superlattice like Fe_3Al;
β_2 superlattice like FeAl or NiAl;
m magnetic alloys;
n non-magnetic alloys.

The diagram shown in fig. 4 includes the whole of the ternary system. The aluminium-rich corner is, however, not fully described in the present paper. There are a large number of small phase fields having nothing in common with the face-centred cubic and body-centred cubic phase fields which are the main subjects of the present investigation. The evidence for the phase boundaries at the aluminium-poor end of the system is summarized in Tables I–V, and explained in the text. In the two-phase and three-phase alloys the relative proportions of the different constituents were as far as possible estimated from the intensities of the lines.

TABLE I. X-RAY POWDER PHOTOGRAPHS OF FE-NI-AL ALLOYS. FACE-CENTRED CUBIC

Phase fields	Atomic composition			Approximate proportions of phases	Lattice spacings	Radiation	Remarks (see footnotes)
	Fe	Ni	Al				
α	15	80	5	—	3·5428	Cu, Co	*N*
α	22·5	75	2·5	—	3·5469	Cu, Co	*LN*
α	35	60	5	—	3·5700	Cu, Co	*LN*
α	64	32	4	—	3·5856	Cu, Co	*N*
α	57	38	5	—	3·5886	Cu, Co	*LN*
α_1	5	85	10	—	3·5404	Cu, Co	*LS*, *a*
$\alpha_1+\alpha_1'$	10	80	10	α_1 6 α_1' 1	3·5435	Cu, Co	*FF*, *SS*
$\alpha_1+\alpha_1'$	12·5	75	12·5	a_1 2 α_1' 3	—	Cu, Co	*FF*, *LSS*, *b*
$\alpha_1+\alpha_1'$	20	70	10	α_1 3 α_1' 2	—	Cu	*FF*
α_1	25	70	5	—	3·5573	Cu, Co	*LS*
α_1	25	67·5	7·5	—	—	Cu, Co	*S*
α_1	17·5	70	12·5	—	3·5621	Cu	*S*, *a*
α_1	5	75	20	—	3·5654	Cu	*S*
α_1	10	70	20	—	3·5736	Cu, Co	*S*, *b*
α_1	15	65	20	—	—	Cu, Co	*S*
α_1	30	60	10	—	3·5750	Cu, Co	*S*
α_1	40	50	10	—	—	Co	*S*

L, a photograph was taken in a larger camera.
S, superlattice lines were observed, which were stronger with Co radiation than with Cu.
SS, two superlattice patterns were observed.
N, no superlattice lines were visible with either copper or cobalt radiation.
FF, two face-centred cubic patterns were observed.
$\alpha_1+\alpha_1'$, in this two-phase area, α_1' is used to distinguish the constituent with the larger lattice spacing and greater Al content.
a, lump-annealed at 1200° C.
b, lump-annealed at 1250° C.

THE FACE-CENTRED CUBIC α-PHASE FIELD

The face-centred cubic α-phase field of fig. 4 includes the face-centred cubic α-alloys of the nickel-aluminium system, as well as the face-centred cubic iron-nickel alloys (usually called γ). After slow cooling, the binary nickel-aluminium alloys form two separate face-centred cubic phases. In the nickel-rich phase the structure is random; Ni_3Al has an ordered structure, with the nickel atoms at face centres and the aluminium atoms at cube corners. In the ternary α-solid solution, these phase fields are both linked up to the face-centred cubic iron-nickel alloys. They must therefore be regarded as parts of a single-phase field, with a miscibility gap, as in the phenol-water-acetone system.

In the binary nickel-aluminium system, the miscibility gap corresponds to a change from the disordered structure (α), to the ordered structure (α_1) of Ni_3Al. In the ternary system this is no longer the case. On either side of the gap, the alloys are ordered, nickel tending to go into face centres. In the two-phase area the two constituents differ in composition only. For convenience the aluminium-rich constituent is called α_1', to distinguish it from the aluminium-poor constituent α_1. The two constituents can be recognized on the X-ray photographs by their different lattice spacings. The dimensions increase almost proportionately to the amount of added aluminium, with a break corresponding to the miscibility gap. Duplex alloys are characterized by a doubling of the high-order reflexions, including the superlattice lines.

Within the face-centred cubic phase field a broken line is drawn to distinguish the ordered from the disordered structures. This boundary is of course only valid for a particular heat treatment. Under the conditions of the experiment, the ordered structures in the α-phase field do not include the composition $FeNi_3$. At least 4 atomic % of aluminium is needed to give a superlattice. To extend the range of ordered structures, it will probably be necessary to heat the alloys for long periods at temperatures below 600° C. It may then be possible to form a superlattice at the composition Ni_3Fe.

Least aluminium is needed when the nickel content approaches 75 %. Here the structure may be regarded as Ni_3Fe in which a little iron is replaced by aluminium at cube corners, all face centres being filled by nickel. Iron and aluminium are interchangeable at cube corners, Ni_3Al having just the same type of structure as Ni_3Fe. Along the line Ni_3Al–Ni_3Fe the superlattice is of the general type Ni_3X, where X is either Al or Fe. In corroboration, the α_1 superlattice lines are comparatively strong with cobalt radiation, which accentuates the difference in scattering power between nickel and iron, and weak or absent with copper radiation which reduces this difference.

The same type of superlattice extends far beyond the line of compositions where the nickel content is exactly 75 atomic %. The nickel atoms at face centres may be replaced by iron and the iron or aluminium atoms at cube corners may be replaced by nickel. In general, the distinction between nickel and aluminium atoms is much more rigorous than between nickel and iron. Nickel keeps as far as possible to face centres; aluminium as far as possible fills cube corners. The iron atoms are more accommodating, making up the balance on either side as required. This rule is probably valid throughout the whole range of the α_1 structures.

TABLE II. X-RAY POWDER PHOTOGRAPHS OF FE-NI-AL ALLOYS. FACE-CENTRED CUBIC AND BODY-CENTRED CUBIC

Phase fields	Atomic composition Fe	Ni	Al	Approximate proportions of phases		Lattice spacings	Radiation	Remarks (see footnotes)
$\alpha+\beta$	65	30	5	α	200	3·5841	Cu, Co	*N*
				β	1			
$\alpha+\beta$*	75	20	5	α	4	3·5815	Cu, Co	*N*
				β	1			*N*
$\alpha+\beta$	90	7·5	2·5	α	1			*N*
				β	10	2·8618	Co	*N*
$\alpha+\beta_2$	60	32·5	7·5	α	25	3·5821	Cu, Co	*LN*
				β_2	1			*S*
$\alpha+\beta_2$	52·5	35	12·5	α	3	3·5858	Cu, Co	*N*
				β_2	1			
$\alpha_1+\beta_2$	50	40	10	α_1	20	3·5857	Cu	*S*
				β_2	1	2·8654	Co	*S*
$\alpha_1+\beta_2$	30	55	15	α_1	75	3·5823	Cu, Co	*S*
				β_2	1			*S*
$\alpha_1+\beta_2$	20	60	20	α_1	20	3·5813	Cu	*S*
				β_2	1	2·8658	Co	*S*
$\alpha_1+\beta_2$	5	70	25	α_1	200	3·5772	Cu, Co	*S*
				β_2	1			*S*
$\alpha_1+\beta_2$	25	50	25	α_1	2	3·5850	Co	*S*
				β_2	5	2·8658		*S*
$\alpha_1+\beta_2$	15	55	30	α_1	1			*S*
				β_2	7	2·8638	Co	*S*
$\alpha_1+\beta_2$	12·5	55	32·5	α_1	1			*S*
				β_2	50	2·8650	Co	*S*

L, S, N have the same meaning as in Table I.

* The lattice spacing shows that the alloy is on the boundary of the three-phase field.

The lattice spacings of the α-phase alloys are least for pure nickel. The increase on addition of iron and aluminium follows on the same general lines as in the binary systems iron-nickel and nickel-aluminium. There are maxima near Ni_3Al and not far from 40 atomic % of nickel in the binary nickel-iron system. The corresponding maxima in the ternary system are only slightly displaced. The rise in spacing from pure nickel to the two maxima is fairly regular, but beyond these points there is a sharp drop to the phase boundary. This irregularity makes it difficult to use spacing measurements to fix the phase boundaries.

TABLE III. X-RAY POWDER PHOTOGRAPHS OF FE-NI-AL ALLOYS.
BODY-CENTRED CUBIC

Phase fields	Atomic composition Fe	Ni	Al	Position on diagram	Lattice spacings	Radiation	Remarks (see footnotes)
Boundary	50	20	30	On boundary of	2·8834	Co	*S*, *F*
of β_2	62·5	12·5	25	$\beta+\beta_2$ field	2·8820	Co	*S*, *a*
β_2	75	5	20	Near $\beta+\beta_2$	2·8845	Co	*SF*, *b*
β_2	68·75	6·25	25	Near $\beta+\beta_2$	2·8835	Co	*SF*
β_2	55	15	30	Near $\beta+\beta_2$	2·8868	Co	*S*
β_2	68·4	5·3	26·3	Near β_1	2·8893	Co	*SF*
β_1	65	5	30		2·8884	Co	S_1F
β_1	70	2·5	27·5		2·8896	Co	S_1F
β_1	72·5	2·5	25			Co	S_1F
β_2	47·5	20	32·5	Near $\beta+\beta_2$	2·8861	Co	*S*
β_2	35	30	35	Near $\beta+\beta_2$	2·8804	Co	*S*
β_2	30	35	35	Near $\beta+\beta_2$	2·8795	Co	*S*
β_2	25	45	30	Near $\alpha_1+\beta_2$	2·8691	Co	*S*
β_2	17·5	50	32·5	Near $\alpha_1+\beta_2$	2·8671	Co	*S*
β_2	10	55	35	Near $\alpha_1+\beta_2$	2·8658	Co	*S*
β_2	5	57·5	37·5	Near $\alpha_1+\beta_2$	2·8649	Co	*S*

S indicates a superlattice of the β_2 type (like FeAl or NiAl).

S_1 indicates a superlattice of the β_1 type (like Fe_3Al).

F indicates the presence of a small amount (up to 5 %) of a face-centred cubic pattern which is also found mixed with Fe_3Al. It is isomorphous with a structure found in the Fe-Al-C system and with Fe_4N. It only appears when pure Al is used. Commercial aluminium suppresses this phase. It cannot be represented on the equilibrium diagram, and never exceeds about 5 % of the total.

a, lump-annealed at 1400° C.

b, lump-annealed at 1250° C.

With the heat treatment given in the present series of experiments, the limits of the α-phase were fixed from a series of powder photographs showing only a trace of the body-centred cubic pattern. With the 9 cm. camera, it is possible to detect about 1 part in 300 of the body-centred cubic phase, in equilibrium with the face-centred cubic phase.

All the ternary face-centred cubic alloys which we have examined are ferromagnetic. There are however, in all probability, two small non-magnetic regions in this phase field. One borders on Ni_3Al, the other touches the three-phase field, $\alpha+\beta+\beta_2$. This can be deduced from the following facts. The binary alloys near Ni_3Al are non-magnetic. It seems probable that neighbouring ternary alloys will behave in a similar manner, though the nearest ternary alloys which we have examined (e.g. $FeNi_{15}Al_4$) are strongly ferromagnetic. That the face-centred cubic constituent of the three-phase $\alpha+\beta+\beta_2$ alloys is non-magnetic may be deduced from the behaviour of the three-phase alloys richest in this constituent. Both $Fe_{25}Ni_{12}Al_3$ and Fe_6Ni_3Al are scarcely magnetic after cooling to room temperatures. On the other hand, the neighbouring single-phase α-alloy $Fe_{16}Ni_8Al$ and the two-phase alloys $Fe_{13}Ni_6Al$ and $Fe_{24}Ni_{13}Al_3$ are strongly ferromagnetic. The magnetic boundary is therefore at the tip of the α-phase, where it touches the three-phase triangle.

THE BODY-CENTRED CUBIC β-PHASE FIELD

In fig. 4 the body-centred cubic β-phase field consists of a wide area extending on either side of the line FeAl–NiAl, together with a narrow tongue stretching along the binary iron-aluminium edge of the diagram to pure iron. The two main areas constitute the β and β_2 regions, which are partially separated by the β_1 area. The three regions differ only in atomic distribution. β has the atoms at random; in β_2 cube centres differ from cube corners; in β_1 alternate cube corners are different. The β_2 area is further subdivided into a ferromagnetic region m and a non-magnetic region n. All β and β_1 alloys are ferromagnetic at room temperature.

The greater portion of the body-centred cubic phase field belongs to the β_2 region. The ideal formula for this type of structure is XAl, where X is either iron or nickel. There are no limits to the miscibility of FeAl and NiAl. For example, Fe_3NiAl_4 and $FeNiAl_2$ and $FeNi_3Al_4$ are all solid solutions in which aluminium fills cube corners and cube centres are taken up by a mixture of iron and nickel atoms. There is no further regularity in the distribution of the iron and nickel atoms, at least in the slowly cooled state.

This simple type of solid solution is of course only possible when there are exactly enough aluminium atoms. When there are fewer than 50 % of aluminium atoms iron and nickel are found to behave in a dissimilar manner. The missing aluminium atoms at cube corners are replaced by iron rather than by nickel. As far as the atomic proportions will permit, nickel goes to cube centres, while aluminium occupies cube corners. The iron atoms are distributed between centre and corner positions to make up the correct atomic composition. As far as we know, at present, these alloys (with less than 50 atomic % of aluminium) have two atoms per unit cell, but we have not yet made density measurements on all the ternary alloys.

When there are more than 50 % of aluminium atoms, the ternary alloys probably behave like the binary nickel-aluminium alloys with excess of aluminium. The aluminium atoms are then confined to cube corners, the deficiency of nickel atoms being no longer replaced by aluminium in cube centres. For the most part, iron-nickel-aluminium alloys with more than 50 atomic % of aluminium also appear to have less than two atoms per unit cell. The lattice spacing drops steeply as the deficiency increases, just as in the binary nickel-aluminium system.

The lattice spacings of the β-alloys behave in a very irregular fashion, and are unreliable for fixing the phase boundaries. There is scarcely any range of compositions over which a linear relationship exists. The maximum lattice spacing is found at the composition FeAl. Thence it falls first steeply and then gradually to NiAl. The line FeAl-NiAl corresponds approximately with a ridge of high spacing values. With much more or less than 50 atomic % of aluminium, the spacing drops. Space is lacking to give a full description of all the complicated changes in lattice spacing which occur in the β phase. A more detailed account will be published later.

In the β_2-phase field, the magnetic boundary at room temperatures follows a remarkable curve. Starting near Fe_2Al, which happens to be a point of minimum lattice spacing in the iron-aluminium system, the limit of ferromagnetism at first moves towards NiAl. The aluminium-rich alloys are naturally the non-magnetic ones. Before reaching the nickel-aluminium binary edge of the diagram, the magnetic boundary swings round at right angles, moving in the direction of pure nickel. Here the iron-rich alloys are ferromagnetic, those containing little iron being paramagnetic only, as far as we could tell from a cursory examination.

β_1 alloys are related to Fe_3Al. In the ideal superlattice, cube centres and half the cube corners are iron. The aluminium atoms go in alternate cube corners. In the binary alloys, it is possible to add more aluminium to this structure, replacing more of the iron atoms at cube corners. In the ternary alloys, nickel also replaces some of the iron atoms. For example, $Fe_{28}NiAl_{11}$ and $Fe_{13}NiAl_6$ both have the Fe_3Al type of structure after slow cooling. With a little more nickel, the structure changes to the β_2 type, like FeAl, where the only distinction is between cube centres and cube corners as a whole. For example, $Fe_{13}NiAl_5$ has this type of superlattice.

In this part of the system, both binary FeAl alloys and ternary FeNiAl alloys exhibit an unexpected feature. Small traces of another constituent, not shown on the phase diagram, have been found whenever the composition lies close to Fe_3Al. The amount of this constituent varies with the purity of the aluminium used for making the alloys. Contrary to expectation, purer aluminium gives more of the extra phase. In a former investigation (Bradley and Jay 1932), when less pure aluminium was used, only the normal body-centred cubic pattern was found. The extra phase is face-centred cubic, with a lattice spacing about 3·7 A. In this it resembles Fe_4N or Fe_3AlC. There is, however, no reason to suppose that the alloys contain an appreciable quantity of either nitrogen or carbon, so that there must be another explanation for the occurrence of the phase.

TABLE IV. X-RAY POWDER PHOTOGRAPHS OF FE-NI-AL ALLOYS. TWO BODY-CENTRED CUBIC LATTICES IN EQUILIBRIUM

Phase field	Atomic composition			Approximate proportion of phases	Lattice spacings	Radiation	Remarks (see footnotes)
	Fe	Ni	Al				
$\beta+\beta_2$	35	35	30	β 1		Co	*D*
				β_2 10	2·8770		*S*
$\beta+\beta_2$	50	27·5	22·5	β 2		Co	*D*
				β_2 3			*S*
$\beta+\beta_2$	33·3	33·3	33·3	β 1		Co	*D*, *b*
				β_2 50	2·8785		*S*
$\beta+\beta_2$	50	25	25	β 2	2·8680	Co	*D*
				β_2 3	2·8778		*S*
$\beta+\beta_2$	60	20	20	β 3		Co	*D*, *a*
				β_2 2			*S*
$\beta+\beta_2$	75	12·5	12·5	β 6	2·8680	Co	*D*, *b*
				β_2 1	2·8782		*S*
$\beta+\beta_2$	87·5	6·25	6·25	β 10	2·8679	Co	
				β_2 1			*S**
$\beta+\beta_2$	90	5	5	β 20	2·8655	Co	
				β_2 1			*S**
$\beta+\beta_2$	45	25	30	Doubtful		Co	*D*, *S*
$\beta+\beta_2$	81·25	6·25	12·5	β 3	2·8731	Co	
				β_2 1			*S*
$\beta+\beta_2$	85	5	10	β 5	2·8700	Co	
				β_2 1			*S**
$\beta+\beta_2$	85	2	13	β 20	2·8772	Co	*F*
				β_2 1			*LS*
$\beta+\beta_2$	75	6·25	18·75	β 2	2·8807	Co	*F*
				β_2 1			*S*

L, *S*, as in Table I.
F, as in Table III.
D, the α-doublets of 220 and 330 are each split into two components.
*S**, the superlattice lines indicate β_2 in addition to β (see text).
a, lump-annealed at 1250° C.
b, lump-annealed at 1400° C.

ALLOYS CONTAINING TWO BODY-CENTRED CUBIC LATTICES $(\beta+\beta_2)$

A comparison of the new diagram (fig. 4) with Köster's diagram, shown in a simplified form in fig. 3, indicates that the two-phase area found by Köster is to be divided into four separate phase fields. This is possible on account of the remarkable shape of the body-centred cubic phase field. After slow cooling, alloys such as Fe_2NiAl break up into two constituents, one is of the β type, the other of the β_2 type. The two-phase field $\beta+\beta_2$ is really a miscibility gap in the body-centred cubic region.

Köster had already recognized that these alloys had a duplex structure, but he thought that one constituent was face-centred cubic, whereas after slow cooling both are body-centred cubic. It would, of course, have been impossible to establish this fact from a microscopic study alone, unchecked by X-ray evidence. From our powder photographs, we have clear proof that two body-centred cubic lattices are present in many different alloys. For example, the 310 reflexions from Fe_6NiAl, Fe_3NiAl, Fe_2NiAl, FeNiAl and $Fe_7Ni_7Al_6$ all show two sets of $K\alpha$ doublet, corresponding to different lattice spacings. The 220 reflexion is also duplicated, whereas the reflexion from 330 and 221 is a clear-cut $K\alpha$ doublet. This proves that only one constituent (β_2) has a superlattice. The lattice spacing of the β_2 constituent is always greater than that of the β constituent. The intensity of the superlattice lines increases steadily as nickel and aluminium replace iron. At the same time the intensity distribution in the 310 doublet becomes modified, the lines due to the β-phase gradually weakening, while those due to the β_2-phase strengthen. In fact we are dealing with a typical two-phase field.

The greatest difference in spacing between the β and β_2 constituents, which is less than 0·4 %, occurs near Fe_2NiAl. An X-ray investigation of some other $\beta+\beta_2$ alloys is not free from pitfalls. It is not always possible to recognize the presence of the second constituent. Even where the spacing differences are greatest, there is considerable overlapping of the $K\alpha$ doublets. As nickel is replaced by aluminium, the difficulty increases. The lattice spacings of the two phases gradually become more and more nearly equal. Finally there is not the slightest difference between them. It requires the cumulative evidence of all the two-phase alloys to show where the boundary lies on this side.

In the iron-rich alloys, the boundary between the structures without a superlattice (β) and the two-phase $\beta+\beta_2$ area, was recognized by the strength of the superlattice lines belonging to the β_2 constituents. The presence of the superlattice lines proved the existence of the β_2 phase in such alloys as $Fe_{18}NiAl$ and $Fe_{17}NiAl_2$. At first it was thought that these alloys were single-phase, but calculation showed that if that were true the superlattice lines would be too weak to be observed. The following considerations will show why a superlattice line should be stronger in a two-phase alloy than in a single-phase alloy of the same composition.

Let a homogeneous alloy of the β_2 type have a superlattice line of strength s. If it be diluted by addition of a second constituent, the superlattice line will be less intense. Neglecting for simplicity any complications due to absorption, the intensity will be directly proportional to the concentration of the β_2 constituent. With x % of the β_2 constituent the intensity of the superlattice line will be $xs/100$. Now suppose that the diluting phase reacts with the β_2 constituent to form a single-phase alloy with the β_2 type of superlattice. Instead of the *intensity* of the superlattice line being reduced by the ratio $x/100$, the *structure factor* will be reduced in this ratio. (It is supposed that the diluting phase does not contribute to the superlattice after reaction.) Since the intensity is proportional to the square of the structure factor, the intensity reduction is in the proportion $(x/100)^2$. The intensity of the superlattice line is therefore $(x/100)^2 s$ for the single-phase alloy against $xs/100$ for the two-phase alloy.

For example, the alloy $Fe_{18}NiAl$ as a single-phase would have a superlattice intensity not exceeding $(1/10)^2$ that of NiAl. This would be quite invisible. As a two-phase alloy, it might give a maximum superlattice intensity as much as 1/10 that of NiAl. This could easily be observed. In fact, this alloy gives visible superlattice lines and is therefore two-phase. The same is true of the alloy $Fe_{17}NiAl_2$. The intensities of the superlattice lines are such that these compositions must be well removed from the boundary of the two-phase area.

The lattice spacing measurements also give an important clue to the position of the $\beta + \beta_2$ boundary. The spacing measurements within the two-phase field are on the whole consistent with one another, giving the general trend of the tie-lines. Within the single-phase field of the β_2 alloys, the spacings are more erratic. For example $Fe_{13}NiAl_5$ is relatively high compared with $Fe_{11}NiAl_4$. Neither alloy could be conveniently fitted on a tie-line, which indicates that in all probability they lie in the single-phase field.

Using such considerations, we decided that the phase boundary probably ran in the manner shown in fig. 4. It is difficult to be sure whether any given alloy lies on one side of the boundary or the other. The spacing measurements cannot be used directly to make a quantitative estimate of the position of the boundary, on account of the superimposition of β and β_2 lines. Only approximate estimates are possible, and except near the end of the phase field, it is impossible to recognize the presence of the β-constituent, as there is nothing to distinguish the β lines from the β_2 lines.

$Fe_{12}NiAl_3$ and $Fe_{85}Ni_2Al_{13}$ are probably in the two-phase field. The superlattice intensities agree with this supposition. It is unlikely that the latter would have a superlattice if it were a single-phase alloy, since the composition is little removed from the binary iron-aluminium alloys, which in this region are disordered. $Fe_{12}NiAl_3$ has quite a strong superlattice after slow cooling, which is considerably weakened after quenching from 900° C. This is also consistent with a position in the two-phase field, if it is supposed that the proportion of β_2 to β is less at high temperatures.

TABLE V. X-RAY POWDER PHOTOGRAPHS OF FE-NI-AL ALLOYS. TWO BODY-CENTRED CUBIC LATTICES AND A FACE-CENTRED CUBIC LATTICE

Phase field	Atomic composition Fe	Ni	Al	Approximate proportions of phases		Lattice spacings	Radiation	Remarks (see footnotes)
$\alpha+\beta+\beta_2$	62·5	30	7·5	α	50		Co	
				β	1			*D*
				β_2	3			*S*
$\alpha+\beta+\beta_2$	60	30	10	α	5	3·5815	Cu, Co	
				β, β_2	1	$2{\cdot}867_6$		*D* *S*
$\alpha+\beta+\beta_2$	55	30	15	α	5		Co	
				β	2	$2{\cdot}868_6$		*D*
				β_2	1			*S*
$\alpha+\beta+\beta_2$	50	30	20	α	1		Co	
				β	2	$2{\cdot}862_4$		*D*
				β_2	3	$2{\cdot}872_2$		*S*
$\alpha+\beta+\beta_2$	85	10	5	α	1		Co	
				β	100	2·8641		
				β_2	10			*S**
$\alpha+\beta+\beta_2$	30	40	30	α_1	1		Co	Doubtful
				β				
				β_2	200	$2{\cdot}872_4$		*S*

S, superlattice lines were observed.
D, *S**, as in Table III.

FeNiAl and $Fe_9Ni_5Al_6$ are definitely two-phase, as is shown by the 310 α-doublet. There are two alloys which have unique magnetic properties, showing strong coercivity after slow cooling. These are $Fe_5Ni_2Al_3$ and Fe_5NiAl_2. It is possible that they lie exactly on the phase boundary.

THE THREE-PHASE AREA $\alpha+\beta+\beta_2$

The existence of the $\beta+\beta_2$ area leads logically to a three-phase region $\alpha+\beta+\beta_2$. In this area the composition of each phase is invariant, provided that the temperature is kept constant. These compositions are given by the corners of the three-phase triangle, and are very susceptible to change of temperature. It is beyond the scope of the present work to give more than a hint of the direction in which the boundaries move as the temperature is raised.

After slow cooling, the three-phase field includes at least five of the alloys which we have examined. It is more difficult to distinguish the two body-centred cubic constituents than in some of the duplex alloys. The lattice spacings are closer together here than in the case of Fe_2NiAl. Nevertheless, there is no doubt of the two body-centred cubic lattices in such an alloy as $Fe_5Ni_3Al_2$.

The three corners of the triangle are best fixed from alloys containing a large amount of one constituent and only a little of the other two. For example, $Fe_{36}Ni_3Al$ containing only a little of the α and β_2 phases gives the approximate position of the corner of the triangle which touches the β-phase field. Fe_6Ni_3Al is nearly pure α; $Fe_3Ni_4Al_3$ is almost pure β_2. The corners of the three-phase triangle are probably at $Fe_{62}Ni_{31}Al_7$, $Fe_{90}Ni_7Al_3$ and $Fe_{28}Ni_{41}Al_{31}$.

After quenching from 800° C. different compositions were found for the three phases in equilibrium. The composition of the α-constituent is displaced towards the centre of the diagram; β and β_2 probably come together. It is not improbable that finally a temperature is reached where the compositions of the β and β_2 phases coincide, so that the three-phase triangle disappears. The equilibrium will then be as shown in fig. 2.

Each edge of the three-phase triangle is a tie-line. Any alloy lying on this line consists of a mixture of the two phases whose compositions are given by the adjacent corners of the triangle. Every two-phase field is covered by a family of such tie-lines. At first, in the neighbourhood of the three-phase triangle, the tie-lines run almost parallel to the corresponding edge of the triangle. Later they may change direction. For example, in the $\alpha+\beta$ area, the tie-lines turn through a small angle and finally run parallel to the iron-nickel edge of the diagram. The change in direction is much more marked in the case of $\alpha_1+\beta_2$ phase field where the tie-lines finally swing round to the direction of the nickel-aluminium edge of the diagram. In the $\beta+\beta_2$ area, the tie-lines remain throughout roughly parallel to the corresponding edge of the three-phase triangle.

THE ALUMINIUM-RICH ALLOYS

Only a brief mention of the aluminium-rich alloys can be given here. There are seven small single-phase fields, separated by alternate two- and three-phase fields. Two of the phases occur in the binary iron-aluminium system and have practically no extension in the ternary diagram. They may be represented by the formulae $FeAl_2$ and Fe_2Al_5. A third phase which occurs in the iron-aluminium system stretches right into the interior of the ternary diagram. This includes the composition Fe_2Al_7 (not $FeAl_3$); here about one-third of the iron atoms are replaceable by nickel.

Of the phases occurring in the nickel-aluminium system, Ni_2Al_3 although permitting considerable variation in the aluminium content, only dissolves about 2 % of iron. On the other hand, $NiAl_3$, permitting practically no variation in the aluminium content, yet allows about one-fifth of the nickel atoms to be replaced by iron.

One of the most remarkable features of the system is the isomorphism of the ternary compounds Fe_3NiAl_{10} and $FeNiAl_9$ with structures found in the binary cobalt-aluminium system. A complete account of this system will be published shortly. At present we can only state that Fe_3NiAl_{10} has exactly the same type of structure as Co_2Al_5, while $FeNiAl_9$ has the structure of Co_2Al_9. These facts indicate a parallel behaviour between cobalt-aluminium alloys and ternary alloys containing iron and nickel in certain proportions.

The structures of the two nickel-aluminium phases, Ni_2Al_3 and $NiAl_3$, have been described elsewhere (Bradley and Taylor 1937*a*). That of Co_2Al_5 will be published shortly. The relationship of Fe_3NiAl_{10} to the cobalt-aluminium alloy follows directly from the atomic grouping. That of $FeNiAl_9$ to Co_2Al_9 must await a solution of these structures. The three complicated structures corresponding to the iron-aluminium binary alloys still await solution.

ACKNOWLEDGEMENTS

We are deeply indebted to Professor W. L. Bragg, F.R.S., for his great interest in the work which was commenced in the Physical Laboratories of the University of Manchester and completed at the National Physical Laboratory. One of us (A. T.) was in receipt of a grant from the Permanent Magnet Association, for the purpose of carrying out this research which is being continued in co-operation with the British Electrical Research Association.

SUMMARY

X-ray powder photographs of slowly cooled alloys of iron, nickel and aluminium have been used to construct a phase diagram for all compositions. The system falls into two distinct portions. Up to 50 atomic % of aluminium there are only body-centred cubic and face-centred cubic structures. Beyond 50 % of aluminium, the diagram is extremely complex and will be described in a later paper.

The face-centred cubic phase field includes superlattice structures related to Ni_3Al or Ni_3Fe (α_1), as well as disordered structures (α). It encloses a miscibility gap where two face-centred cubic superlattice phases (α_1 and α_1') are in equilibrium. As the iron content increases, the gap closes.

The body-centred cubic phase field includes superlattice structures related to Fe_3Al (β_1) and FeAl or NiAl (β_2). There is a two-phase field where the iron-rich β phase (without a superlattice) is in equilibrium with the β_2 phase. In consequence, the system is much more complicated than had previously been supposed. Instead of a two-phase field separating the face-centred cubic and body-centred cubic areas, there are four separate areas, comprising three two-phase fields and a three-phase field. In the latter a face-centred cubic structure (α) is in equilibrium with two body-centred cubic structures (β and β_2). The two-phase fields are $\alpha+\beta$, $\alpha+\beta_2$ and $\beta+\beta_2$ respectively. The corners of the three-phase triangle are placed at the following compositions: α, 62, 31, 7; β, 90, 7, 3; β_2, 28, 41, 31. (The numbers represent atomic % iron, nickel and aluminium respectively.)

REFERENCES

Ageew, N. W. and Vher, O. I. 1930 *J. Inst. Met.* **44**, 83.
Alexander, W. O. and Vaughan, N. B. 1937 *J. Inst. Met.* **61**, 247.
Bradley, A. J. and Jay, A. H. 1932*a* *Proc. Roy. Soc.* A, **136**, 210.
— — 1932*b* *J. Iron Steel Inst.* **125**, 339.
Bradley, A. J., Jay, A. H. and Taylor, A. 1937 *Phil. Mag.* **23**, 545.
Bradley, A. J. and Taylor, A. 1937*a* *Proc. Roy. Soc.* A, **159**, 56.
— — 1937*b* *Phil. Mag.* **23**, 1049.
— — 1937*c* *Nature, Lond.*, **140**, 1012.
Gwyer, A. C. G. 1908 *Z. anorg. Chem.* **57**, 133.
Gwyer, A. C. G. and Phillips, H. W. 1927 *J. Inst. Met.* **35**, 29.
Jette, E. R. and Foote, F. 1936 *Amer. Inst. Min. Engrs*, **120**, 259.
Kiuti, S. 1937 *Journ. Aero. Res. Inst. Tokyo*, **152**, 179 (in Japanese).
Köster, W. 1934 *Arch. Eisenhüttenw.* **7**, 257.
Merica 1930 *National Metals Handbook*, p. 607.
Nishimura, H. and Watanabe, S. 1936 *Suiyokwai Shi*, **9**, 153 (in Japanese).
Owen, E. A. and Yates, E. L. 1937 *Proc. Phys. Soc.* **49**, 178, 307.
Owen, E. A., Yates, E. L. and Sully, A. H. 1937 *Proc. Phys. Soc.* **49**, 315.
Sykes, C. and Evans, H. 1935 *J. Iron Steel Inst.* **131**, 225.

Atomic rearrangement process in the copper-gold alloy Cu_3Au. II

By F. W. Jones, B.Sc., and C. Sykes, D.Sc., Ph.D.

(*Communicated by W. L. Bragg, F.R.S.—Received* 24 *February* 1938)

[Plate 15]

1. Introduction

In a recent investigation (Sykes and Jones 1936) of the properties of the alloy Cu_3Au we described the effect of the ordering process on the energy content of the alloy. The observed behaviour was found to be more complicated than was anticipated from the theoretical work (Bragg and Williams 1934; Peierls 1936), and the existence of a variety of metastable states of an unexpected type was established. The face-centred lattice of Cu_3Au may be considered in terms of the four interpenetrating simple cubic lattices to which it is equivalent. When completely ordered one lattice contains nothing but gold atoms and the other three, copper atoms. As the alloy cools through the critical temperature gold atoms may segregate on any of the four lattices, consequently any individual crystal in the alloy will contain a large number of small volumes or nuclei, each of which has its own consistent scheme of order, but different nuclei will be out of phase with one another. The various metastable states observed experimentally are associated with the presence of these nuclei, since their rate of growth, which determines the rate of disappearance of disordered material from the boundaries, is small compared with the time of relaxation of the material inside the nuclei. The size of the nuclei can be obtained from the width of the superlattice lines present in the X-ray powder diagram; this information, together with a number of specific-heat temperature curves, was used to determine the effect of the nuclei on the energy content of the alloy. It was shown that the experimental results were consistent with the assumption that the nuclei in the individual crystals were separated by narrow boundaries (from about one to two atomic distances in width), the atoms inside the nuclei having the equilibrium degree of order at the temperature of the experiment.

The electrical resistance of an alloy in equilibrium depends on the degree of order; the presence of nuclei corresponding to a metastable state introduces an extra resistance which is a function of the size of nuclei. The present paper describes experimental work carried out to determine the relation between the size of the nuclei and electrical resistance. The results show that at any given temperature the extra electrical resistance is inversely proportional to the size of the nuclei, that is to the number of boundaries, provided the linear dimensions of the nuclei are more than 20 atomic distances. The quantitative relations between nuclei size on the one hand, and resistance, lattice spacing, and intensity of superlattice lines on the other, are consistent with the view that the boundaries are quite narrow. By cold working the material it has been found possible to produce specimens in which the sizes of both nuclei and crystals are of the same order. The rate of growth of nuclei in the cold-worked material is appreciably slower than in the annealed material although the resistance for a given size of the nuclei is almost the same. Using the well-known expressions given by the wave theory for the resistance of a metal (Mott and Jones 1936, p. 268) the probability of reflexion of the conduction electrons at the boundary of the nuclei has been deduced from the experimental results and is about one-tenth (assuming one conduction electron per atom).

2. EXPERIMENTAL WORK

(*a*) *Preparation of the specimens*

All the material was taken from a single ingot of the alloy melted *in vacuo* by Messrs Johnson Matthey. It was cold drawn into one coil of wire 1 mm. diameter and analyses made from samples taken from both ends of the coil. The results were 50·84 % Au and 50·82 % Au, the difference being of the order of the analytical error. The mean analysis 50·83 % Au by weight is 25·0 % Au by atoms. At 1 mm. diameter the wire was annealed for 4 hr. at 950° C. to remove coring. X-ray examination, using the K_α radiation of nickel which allows one doublet to be examined under a high degree of resolution, showed that coring had been completely removed. The surface of the wire was rubbed with emery paper to remove a thin layer which had become slightly impoverished of copper by evaporation, and the coil then cold drawn down to 0·4 mm. diameter, the size used for experimental work.

Since sharp X-ray powder diagrams suitable for intensity measurements had to be obtained from the wire the effects of cold work (which give rise to diffuse diffraction haloes) had to be removed without producing a crystal size in excess of 10^{-3} cm. (otherwise spotty rings would result). The necessary heat treatment, which was critical, was found by trial and error and consisted in an anneal at 550° C. for 30 min. *in vacuo*. The resulting material referred to later as standard material was very homogeneous: the resistivity did not vary by more than ± 1 part in 1000, and the lattice spacing of the X-ray specimens was uniform within ± 1 part in 20,000, the limit of measurement.

The above procedure, which is rather complicated, was found to be essential and indicates the precautions which must be taken with this alloy, in order to produce samples which give identical values of resistance and lattice spacing.

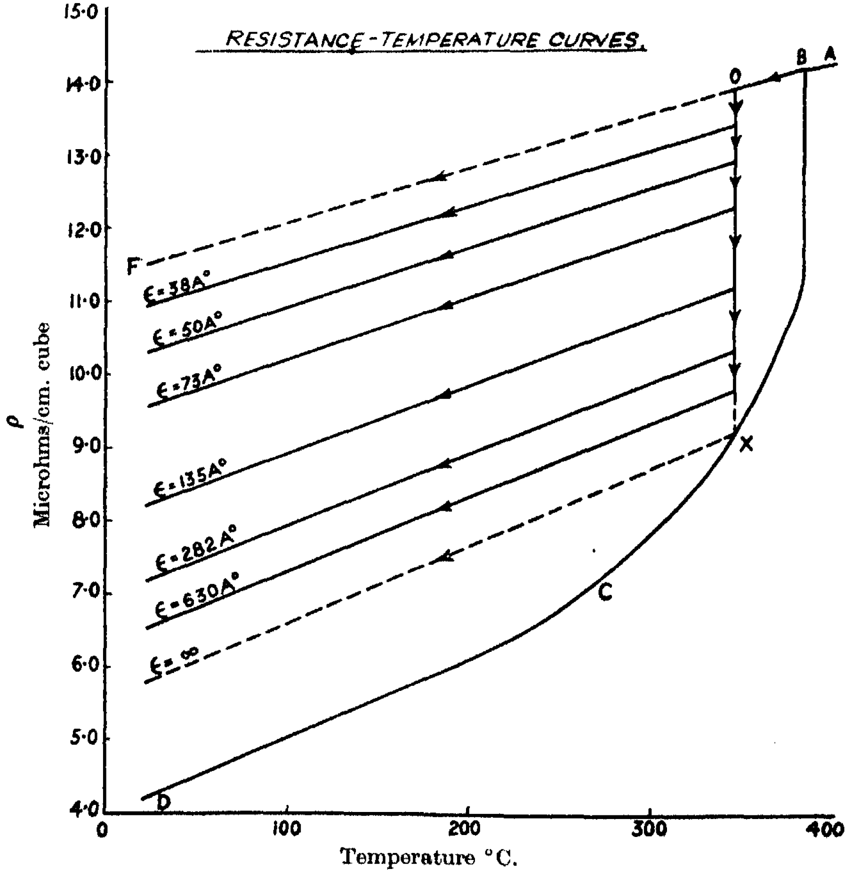


FIG. 1

(b) *Method*

The method of investigation is best described by consideration of the resistance-temperature curves of the alloy. In fig. 1 the curve *ABCD* is the equilibrium resistance temperature curve: *ABF* is the curve obtained if the alloy is quenched in water from a temperature corresponding to the point *A*. At room temperature (*F*) the alloy is completely disordered (see fig. 2, Plate 15). If the quenched specimen is reheated to 350° C. (point *O*) ordering sets in and the resistance gradually falls until it reaches the corresponding point *X* on the equilibrium curve. The mechanism is somewhat as follows: Segregation of gold atoms starts from centres and proceeds until the growing nuclei touch, further diminution in resistance then occurs as certain favoured nuclei grow by absorption of others. Finally, when the nuclei are very large (fig. 5, Plate 15), i.e. when each crystal has its own coherent scheme of order the point *X* is reached. Throughout the annealing treatment the temperature is kept constant and the whole change in resistance is caused by the growth of the nuclei during the atomic rearrangement process. By quenching specimens heat treated at constant temperature for varying periods of time the relation between the size of the nuclei and the electrical resistance can be determined, measurements of both variables being made on the same specimen. (The rate of growth of the nuclei is in all cases so slow that no appreciable change can take place during quenching. At room temperature the alloy is effectively frozen.)

The experimental arrangements were as follows: A number of wires 0·4 mm. diameter and 12·5 cm. long were sealed off in evacuated glass tubes, which were mounted on long glass rods. A furnace, having a uniform temperature zone of suitable length, was heated to the temperature of the experiment, say 346° C., and allowed to become steady, the temperature being controlled by means of a sensitive regulator. An additional furnace was heated to 400° C., the specimens inserted and allowed to come up to temperature, in order that their structure should be completely disordered. They were then withdrawn and after a period of a few seconds inserted into the furnace running at the steady temperature. In this way steady conditions were obtained in a minimum time.

At suitable intervals of time, chosen to produce equal changes in resistance as checked on a pilot specimen in the furnace, specimens were withdrawn and quenched in water. The resistance of the specimens was measured, and then a piece of wire was cut off and the X-ray powder diagram required for the determination of nuclei size was obtained from it. Wire was preferred to powder for the X-ray work throughout, as it enabled both types of measurement to be made on the same specimen.* Experiments were carried out at three different temperatures, 376, 346 and 298° C., and the total annealing periods varied from 13 hr. at 376° C. to 400 hr. at 298° C.

---

* A preliminary experiment showed that nuclei size determinations on wire and powder given the same heat treatment agreed within the limits of experimental error.

(c) *Measurements of nuclei size*

The X-ray photographs were taken in a 9 cm. camera using the K_α radiation of copper. The ordered structure of the alloy Cu_3Au is rich in strong superlattice lines (cf. figs. 3, 4 and 5, Plate 15) and the size of the nuclei can of course be determined from any one of these lines using the formula

$$\epsilon = \frac{\lambda \cos \frac{1}{2}\chi}{\beta},$$

where ϵ is the apparent particle size of the nuclei, referred to subsequently as the nuclei size; λ the wave-length of the radiation used; χ is twice the Bragg angle of the reflexion; and β is the true angular breadth of the line. B_0, the observed breadth of the line, is obtained by integrating the total intensity over the line and dividing by the maximum intensity, this information being obtained by microphotometry from the X-ray film. β is obtained from B_0 by making certain corrections for finite slit width, size of specimen, K_α doublet separation, etc. These various corrections were carried out by the method recently developed (Jones 1938).

The apparent nuclei size obtained from different lines is not the same, in particular the size calculated from the 100 reflexion $\Sigma h^2 = 1$ is appreciably larger than that for the other reflexions. Careful consideration of typical films showed that lines $\Sigma h^2 = 5$ and 6 gave good agreement with lines $\Sigma h^2 = 13$ and 14 for sizes for which reasonably accurate results could be expected from all the lines on the same film. Consequently when the nuclei size was very small (fig. 3) measurements were made on lines $\Sigma h^2 = 5$ and 6, and when it was large (fig. 4) measurements were confined to lines $\Sigma h^2 = 13$ and 14.

The experimental conditions are very favourable for accurate measurements of nuclei size, since the normal breadths of the lines in the absence of broadening, which are required to correct B_0 to β, can be determined directly from the main lattice lines on the same film. In the case of certain specimens examined in the cold-worked condition this was not the case, as the cold working produces a broadening of the main lattice lines as a result of grain refinement and lattice distortion. The "normal" breadth for the cold-worked specimens was assumed equal to that given by the standard wire, as all specimens were the same diameter and were photographed under the same experimental conditions. In this way both the apparent crystal size,

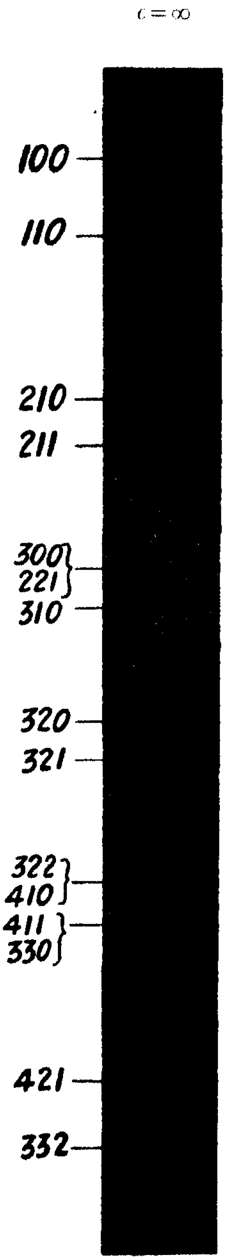


Fig. 5

Fig. 4

Fig. 3

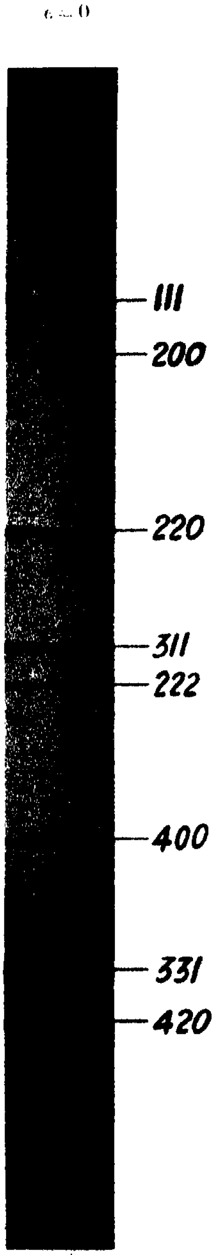


Fig. 2

calculated on the assumption that the whole of the broadening was due to grain refinement, and the nuclei size was determined.

The reproducibility of the measurements between 50 and 250 A for the standard material was found to be ± 5 %. For the cold-worked material, the possible error is greater due to the uncertainty in the correction already mentioned. The absolute accuracy of the determinations of the size of the nuclei will be considered in some detail later in this paper.

3. Experimental results

(*a*) *Effect of nuclei size on resistance*

Fig. 6, curves 1, 2 and 3, shows in graphical form the results obtained at the three different temperatures for the standard material, the resistivity being plotted against $1/\epsilon$, where ϵ is the nuclei size in Angstrom units. The resistivity of the disordered material as quenched is 11·5 ohms/cm. cube,

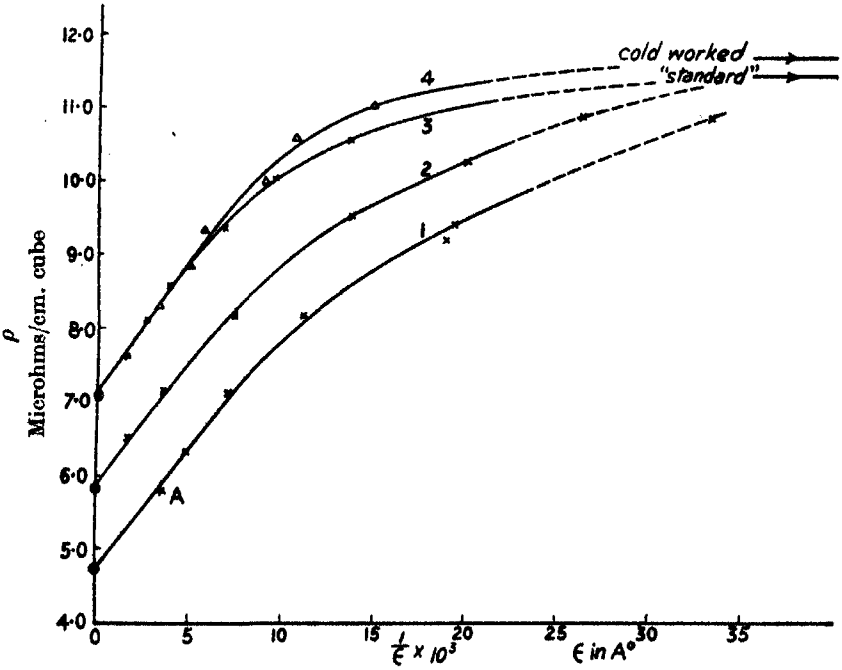


Fig. 6. Specific resistance and nuclei size.

1. 298° C. ×
2. 346° C. ×
3. 376° C. ×

Specimens free from cold work: "standard".

4. 376° C. △ Cold worked.

so that the three curves should approach this value asymptotically as the nuclei size diminishes. For large values of ϵ the experimental points lie on straight lines which have been produced to $1/\epsilon = 0$. The intercept on the resistivity axis should correspond to the resistance of the material in which the size of the nuclei is comparable with the size of the individual crystals of which the specimen is composed, i.e. the equilibrium value. Equilibrium conditions are most rapidly attained by soaking the specimens just below the critical temperature at say 380° C. for a period of 50 hr., during which time the nuclei grow to a size comparable with the crystal size; the specimens can then be cooled fairly rapidly and they still remain in equilibrium. In this way an equilibrium value for 300° C. can be obtained in 100 hr., whereas at least 1000 hr. would be necessary at a constant temperature of 300° C. For example, the point *A* on curve 1, fig. 6, corresponds to an annealing period of 400 hr. at 298° C. and is still a long way from the equilibrium value. The points marked ⊗ corresponding to $1/\epsilon = 0$ were obtained using the special heat treatment process described and are very near to the corresponding points given by the intersection of the lines. (The equilibrium values increase with increasing temperature as the degree of order itself is a function of temperature.)

The measurements at 376° C. were also carried out on cold-worked samples. the cold-worked and standard specimens being treated together. Curve 4 gives the results obtained and is comparable with curve 3, fig. 6. At the commencement of the experiment ($\epsilon = 0$) the resistivity is $11{\cdot}7 \times 10^{-6}$ ohms/cm. cube, i.e. about 2 % more than for the standard material due to the increased resistance associated with lattice distortion. As the annealing proceeds curves 3 and 4 approach each other and the same equilibrium value, and in the region of nuclei size greater than 100 A the two curves coincide, indicating that the resistance is a function only of the nuclei size and is independent of cold-working effects.

(*b*) *Rate of growth of nuclei*

Fig. 7 shows relation between nuclei size and time for the cold-worked and standard materials at 376° C. The rate of growth is very much slower in the cold-worked material, a very unusual result, inasmuch as most metallurgical processes proceed faster in cold-worked than in annealed material. In this particular instance, however, there appears to be no reason to doubt the result shown in fig. 7, since the specimens were made from the same material and heat treated at 376° C. simultaneously.

The behaviour of the standard material at 346 and 298° C. was similar to that shown in fig. 7 for the experiment at 376° C. Various attempts to find a simple relation between the nuclei size and time have been made but were unsuccessful, and we conclude that the rate of growth is not a simple function of either the size of the nuclei or the material in the boundaries.

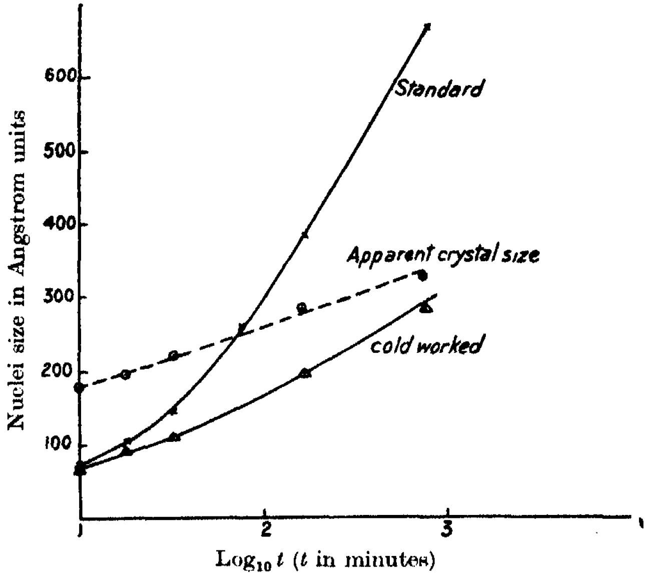


FIG. 7. Growth of nuclei at 376° C.

(c) *Lattice spacing and nuclei size*

Fig. 8 gives the results obtained at 346° C. These are referred to in § 4.

(d) *The effect of nuclei size on the intensity of the superlattice lines*

Although the breadth of the superlattice lines increases with diminishing nuclei size and the lines appear fainter, the integrated intensity falls off relatively slowly. The ratio of intensities of line $\Sigma h^2 = 14$ (a superlattice line) to line $\Sigma h^2 = 16$, a main lattice line, has been determined from the films obtained on specimens heat treated at 376° C. The ratio changes from 1·05 for a nuclei size of 105 A to 1·25 for a nuclei size of 670 A.

4. DISCUSSION OF EXPERIMENTAL RESULTS

The various experimental results recorded in § 3 enable certain speculations to be made as to the behaviour of nuclei and their boundaries. Let us assume that the nuclei have the equilibrium degree of order at the temperature concerned, and that the boundary material is fully disordered; further, that the nuclei are cubic in shape and equal in size.

In our previous paper (Sykes and Jones 1936), using this simplified model, we showed that measurements of energy content were consistent with the conclusion that the boundaries were only two atomic distances wide, viz. 7·5 A, over the range of nuclei size from 40 to 280 A. Measurements quoted in §§ 3 (*c*) and 3 (*d*) are also in agreement with this conclusion.

The intensity of the superlattice lines will be proportional to the amount of ordered material present. On the assumption that the boundaries are 7·5 A wide we have calculated the ratio of ordered material for nuclei 670 A in length to that for nuclei 105 A in length to be 1·19. The ratio of intensities of corresponding superlattice lines is 1·20.

To a first approximation the volume of the unit cell of material containing nuclei will be directly proportional to the relative amounts of ordered and disordered material present. Using the experimental values obtained for the lattice spacings of the alloy in equilibrium at 346° C. and the completely disordered material we have calculated the lattice spacings for various sizes of nuclei, again assuming the boundaries are 7·5 A. wide. The results are shown in fig. 8 marked ⊙, and are in very good agreement with the experimental curve.

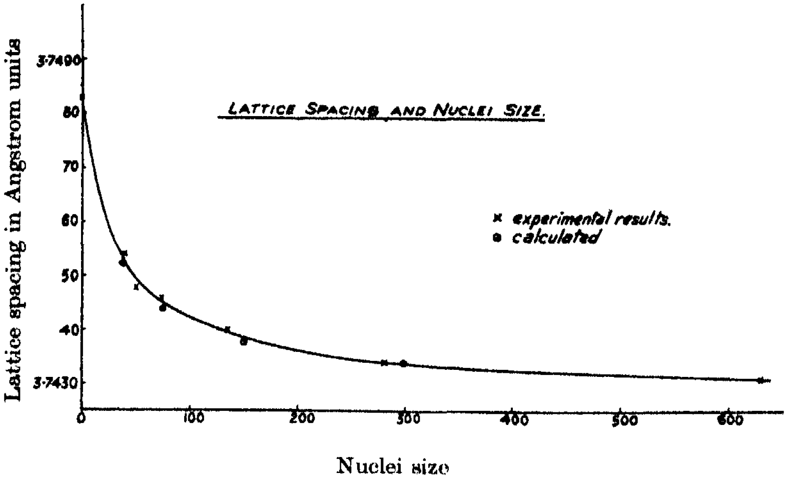


FIG. 8

Whilst the agreement between experimental results and those predicted must be considered to be rather fortuitous in view of the number of simplifying assumptions made, there can be no doubt that the boundaries between nuclei (once the nuclei have attained a size of 40 A or so) are narrow, probably about 7·5 A wide.

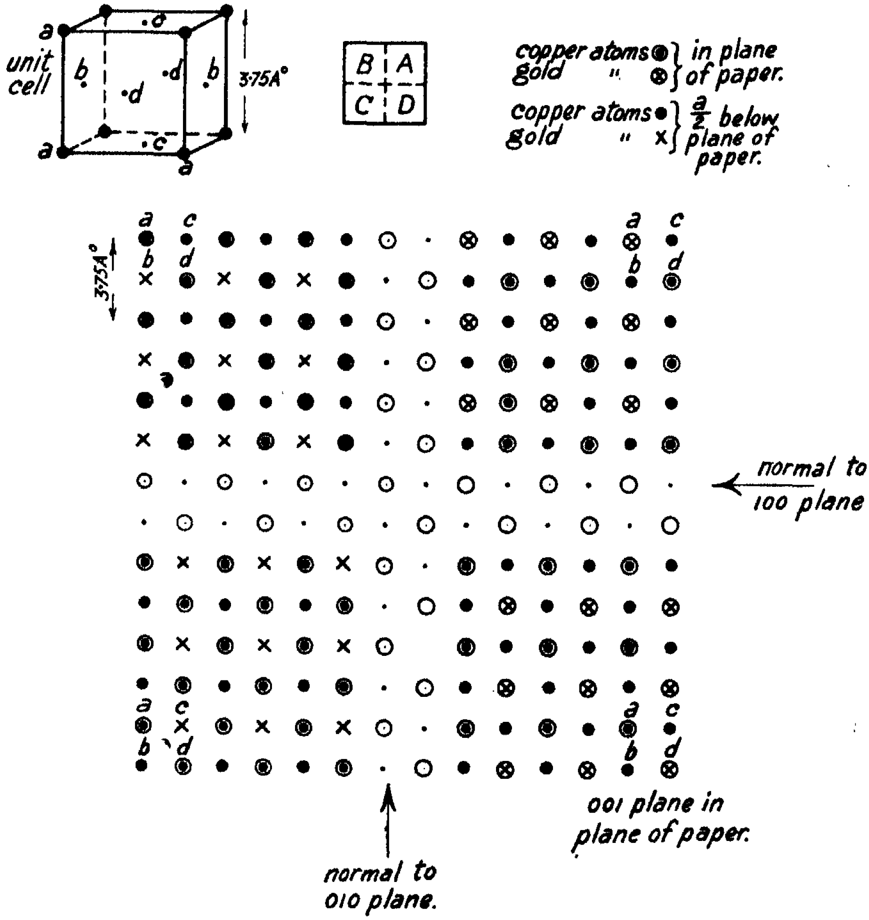
The relation between resistance and nuclei size shown in fig. 6 cannot be satisfactorily explained on the hypothesis that the resistance is directly proportional to the relative amounts of ordered and disordered material present. If this were the case, then for nuclei sizes of 75, 135 and 200 A the width of the boundaries necessary to account for the observed change in resistance would be 11, 40 and 60 A, i.e. the boundaries necessary are much too wide, and further they increase in width with increasing nuclei size which is a most improbable result. Consequently it can be concluded that the extra resistance introduced by the presence of nuclei is not primarily associated with the amount of disordered material present.

The wave theory of metals suggests that the extra resistance may be due to the scattering of the conduction electrons at the boundaries, and the linear relation between resistance and reciprocal nuclei size (see fig. 6) can be interpreted on this basis, since the total scattering will be proportional to the number of boundaries per unit length, i.e. to the reciprocal nuclei size. The effective scattering power of the boundary will diminish for small nuclei sizes, i.e. when the material is more or less disordered, as the distinction between boundary material and ordered material must vanish. Consequently the slope of the curves in fig. 6 should decrease with decreasing nuclei size, which is in agreement with the experimental result.

Reflexion of X-rays from nuclei

If the boundaries are of the order of two atomic distances wide, then some further explanation is required to account for the marked stability of structures containing nuclei, and we have put forward the hypothesis that the various nuclei are antiphase, otherwise they would rapidly coalesce. Dehlinger (1935, p. 154), in discussing the similar case in CuAu, disagrees with this conclusion, his objection being that the intensity of the superlattice lines reflected from conglomerations of antiphase nuclei should be much smaller than the values observed in practice. General considerations would suggest that the energy reflected from nuclei (in phase or antiphase) should be proportional to the total scattering power of the material present. The line breadth and the distribution of energy as a function of the angle of reflexion will vary, of course, with the configuration of the nuclei. As the disordered material will scatter into the general background we expect the intensity of the superlattice lines to be proportional to the amount of ordered material present.

According to our conception of nuclei the arrangement of atoms on the 001 plane is as shown in fig. 9. The four typical sites in the face-centred unit cell are marked *a*, *b*, *c*, *d*, and repetition of each site produces a simple cubic lattice. The four types of nuclei are marked *A*, *B*, *C* and *D*, and only the atoms immediately adjoining the boundaries are considered to be disordered. The behaviour of a linear chain of nuclei is probably similar to that of a number of diffraction gratings of identical spacing joined in such a way that the interval between gratings is an integral number "*n*" times half the spacing. Calculation shows that the interaction term will vanish for a large number of such gratings if the number *n* is ust as likely to be even



a
unit cell
b
d
c
3·75A°
B A
C D
copper atoms ⊛ } in plane
gold " ⊗ } of paper.
copper atoms • } a/2 below
gold " × } plane of paper.
3·75A°
normal to 100 plane
normal to 010 plane.
001 plane in plane of paper.

FIG. 9

as odd, and the intensity of the beam reflected from the composite grating is equal to the sum of the intensities reflected from the individual gratings.

The three-dimensional problem is more complicated and exhibits certain peculiarities. It is clear from fig. 9 that for the 100 reflexion nuclei *A* and *B* are in phase, since alternate planes have the same scattering power; on the other hand, for the 010 reflexion they are out of phase. Similar considerations apply to *C* and *D* and to reflexions from all planes giving rise to superlattice lines.

We have worked out the "interference functions" (Laue 1926*a*, *b*; Jones 1938) for a crystal composed of two nuclei of equal size. From these functions the angular intensity distribution in the reflected X-ray beam can be easily calculated for cases where two Millerian indices are zero, and the results are shown diagrammatically in fig. 10.

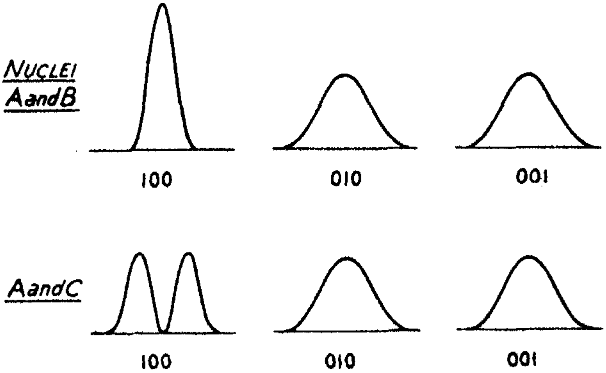


FIG. 10

If the nuclei are in phase in the direction of the line joining their centres (which has been taken perpendicular to the 100 plane), then for the 100 reflexion they behave as a single nucleus of twice the size, for reflexions at right angles they behave as individual units (nuclei *A* and *B*). If they are out of phase in the 100 direction, then for that reflexion they produce zero intensity at the usual maximum, but two adjoining maxima (nuclei *A* and *C*). The intensity reflected per nucleus is the same and independent of whether it acts in phase or out of phase with its neighbours. Thus whilst we have been unable to carry out the calculation for the general case we are of the opinion that the evidence supports the hypothesis that the nuclei are constituted in four different ways in spite of Dehlinger's objection.

Actual size of nuclei

Since adjoining nuclei may act in phase the values of the size of the nuclei obtained by X-ray methods will be larger than the true values. The probability that any nucleus is in a chain n nuclei long is as follows:

Length of chain	1	2	3	4	5	6	7
Probability	0·445	0·296	0·148	0·066	0·027	0·011	0·004

The maximum intensity of the line produced by a chain n nuclei long is n^2 times that of a single nucleus. Using the probability figures given above we find that the maximum intensity of a line reflected from a block of material containing many nuclei of identical size will be 1·98 times that observed if all the nuclei acted independently. The apparent size of the nuclei as determined by X-ray *methods is therefore approximately twice* the true size.* This further strengthens the arguments put forward for the narrow boundary.

Detailed consideration of fig. 6

The wave theory gives the following formula connecting resistivity and the mean free path of the conduction electrons (Mott and Jones 1936, p. 268).

$$\rho = \frac{h}{e^2}\left(\frac{3}{\pi N^2}\right)^{\frac{1}{3}}\frac{1}{\lambda},$$

where ρ is the resistivity, h is Planck's constant, e the electronic charge and N the number of conduction electrons per c.c. Assuming that there is one conduction electron per atom and expressing ρ in microhms per cm. cube and λ in Angstrom units the formula is·

$$\rho = 1{\cdot}45 \times 10^3 \frac{1}{\lambda}.$$

For material where the nuclei size is large compared with the width of the boundaries we may write

$$\frac{1}{\lambda} = \frac{1}{\lambda_0} + \frac{r}{\epsilon},$$

where ϵ is the nuclei size, $1/\epsilon$ the number of boundaries per unit length, and r the probability of an electron being reflected at the boundary. Consequently

---

* There is, of course, an additional factor, in the same direction, which depends on the size distribution (Jones 1938), but this is generally neglected in all measurements of particle size by X-ray methods.

$$\rho = 1{\cdot}45 \times 10^3\left(\frac{1}{\lambda_0}+\frac{r}{\epsilon}\right),$$

the first term giving the resistance of the completely ordered alloy and the second term the additional resistance introduced by the nuclei.

The slope of the curves (fig. 6) allows r to be determined, $d\rho/d\left(\frac{1}{\epsilon}\right)$ is $1{\cdot}65 \times 10^{+2}$ over the straight portion taking into account the fact that the true value of ϵ is half that plotted, and r, the reflexion coefficient, is therefore 0·11. As the resistivity of the completely ordered copper-gold alloy (4×10^{-6} ohms/cm. cube) is roughly twice that expected from a mixture of the constituent metals it is possible that the number of conduction electrons per atom may be appreciably less than in either copper or gold. The value of 0·11 should therefore be regarded as a lower limit. Alternatively it may be concluded that the number of free electrons per atom in the copper-gold alloy is not less than 0·1.

For large nuclei the slopes of the curves plotted in fig. 6 are the same within the limits of experimental error. This suggests either that the resistance of the boundaries is independent of the equilibrium order inside the nuclei, or that the equilibrium degree of order does not vary appreciably in the range of temperature concerned. The first alternative is unlikely. Intensity measurements of superlattice lines obtained using samples in equilibrium indicate that the degree of order in material quenched from just below the critical temperature is about 80 % that of material very slowly cooled to room temperature. The difference in degree of order over the range of temperature 296–376° C. is thus quite small, 10–15 %.

The effect of cold work

The interesting result obtained from the cold-worked specimens is that the relation between resistivity and nuclei size is the same as for the annealed material. This shows in a very clear way that the scattering power of the nuclei boundaries is of quite different order from that of either the grain boundaries or the local irregularities supposed to exist in cold-worked crystals.

The differing rates of growth in the two types of material is not easy to understand. The properties of cold-worked crystals are associated with defects in the regular lattice pattern and nuclei growth across such defects might conceivably be difficult. The strain energy involved in lattice distortion is so small, however (0·1 cal./g.), compared with the ordering energy

(6 cal./g.) that it is doubtful whether an explanation on such lines is forthcoming. Wire drawing produces a fibrous structure (which is partially removed by annealing), and this, if sufficiently enhanced, would account for the difference in behaviour of the two materials, since in the cold-worked specimens growth would be limited to one direction. The values obtained for the apparent particle size in the 100 direction (fig. 7), indicate that this is a possible explanation.

Our thanks are due to the Metropolitan-Vickers Electrical Co. Ltd., for providing the facilities for this work, and in particular we are indebted to Dr A. P. M. Fleming, C.B.E., Director and Manager of Research and Education Departments, for his personal interest in the investigation. Part of the experimental work was carried out in the Physical Laboatories of the University of Manchester, and we thank Professor W. L. Bragg, F.R.S., for his kind interest and encouragement.

SUMMARY

The relation between the size of nuclei and electrical resistance has been determined for the copper-gold alloy Cu_3Au in both the annealed and cold-worked states. The experimental results are interpreted, using the hypothesis that nuclei are antiphase with narrow boundaries. For large nuclei a linear relation, between resistance and the number of boundaries per unit length, is found from which the reflexion coefficient of a single boundary has been deduced. This reflexion coefficient is the same for both cold-worked and annealed material, although the rates of growth of the nuclei in the two cases differ considerably.

REFERENCES

Bragg and Williams 1934 *Proc. Roy. Soc.* A, **145**, 699.
Dehlinger 1935 "Gitteraufbau metallischer Systeme." Leipzig.
Jones 1938 *Proc. Roy. Soc.* A, **166**, 16.
Laue 1926*a* *Z. Krystallogr.* **64**, 115.
Laue 1926*b* *Enzyklopädie der math. Wissenschaften*, **5**, Teil. 3, "Wellenoptik", p. 460.
Mott and Jones 1936 "Properties of Metals and Alloys." Clarendon Press.
Peierls 1936 *Proc. Roy. Soc.* A, **154**, 207.
Sykes and Jones 1936 *Proc. Roy. Soc.* A, **157**, 213.

The statistical mechanics of condensing systems

By Max Born and Klaus Fuchs
Edinburgh

(*Communicated by E. T. Whittaker, F.R.S.—Received* 1 *March* 1938)

Introduction

J. E. Mayer (1937) has published, together with some collaborators, several papers under the same title as the present one.* We consider these papers as a most important contribution to statistical mechanics, and this opinion was shared by the International Conference held in Amsterdam, 26 November 1937, in commemoration of Van der Waals' birth.

One of the present authors gave to this meeting a report on Mayer's work (published in *Physica*, 1937) which was followed by a vigorous discussion on the question as to whether Mayer's explanation of the phenomena of condensation is correct. Doubts about this point were raised by the referee, because it is difficult to comprehend how a method of approximation such as that of Mayer, starting from the gaseous state, can lead to the discontinuity of the density on an isothermal curve which corresponds to condensation. The usual methods for treating the equilibrium of two phases introduce the equation of state of both phases and derive the condition for their co-existence. Mayer's theory does nothing of this kind, but treats all possible molecular arrangements with their proper weight, as if there were only one phase. How can the gas molecules "know" when they have to coagulate to form a liquid or solid? Mayer's mathematical method is too involved to make this point quite clear.

We have devoted a considerable effort to control and clarify these calculations, making ample use of the theory of complex functions, and we believe that we have succeeded in showing rigorously, and in a somewhat simpler way than Mayer himself, that his statements are completely correct. And we have succeeded by our method in going a little farther, in so far as we can derive the conditions for co-existence of several phases (triple point). We have had the privilege of corresponding about these questions with Professor G. E. Uhlenbeck (Utrecht, Holland), to whom we

---

* We have to thank Dr Mayer for sending us the manuscript of another paper (written in collaboration with S. F. Harrison) before publication; this article which has meanwhile appeared (1938) contains the most complete presentation of his theory.

owe several important hints for simplifying the proofs. We publish here the whole theory with complete proofs, so that the reader need not refer to other publications on this subject.*

But we shall confine ourselves to presenting here the general theory as far as it can be developed with the minimum of arbitrary assumptions. The detailed discussion of phenomena in the neighbourhood of the critical point by Mayer and Harrison (1938), based on a special assumption about the irreducible integrals (defined in § 2), will therefore not be considered here.

1. THE PARTITION FUNCTION

We consider a system of equal particles (molecules) which act on one another with central forces.

If $H(p, q) = T(p, q) + V(q)$ is the energy of the system, T and V being its kinetic and potential part, the thermodynamic properties can be derived from the partition function

$$Q = \frac{1}{N!\,h^{3N}}\iint\cdots\int e^{-H(p,q)/kT}\,dp_1\ldots dq_{3N} = \left(\frac{2\pi mkT}{h^2}\right)^{\frac{3N}{2}}\frac{Q_\tau}{N!}, \tag{1·1}$$

where
$$Q_\tau = \iint\cdots\int e^{-V(q)/kT}\,d\tau_1\ldots d\tau_N, \quad (d\tau_n = dx_n\,dy_n\,dz_n). \tag{1·2}$$

The assumption of central forces, $V(q) = \sum_{1\leqslant j<i\leqslant N} v(r_{ij})$, leads to

$$Q_\tau = \iint\cdots\int_{N\geqslant i>j\geqslant 1}\prod\{1+f(r_{ij})\}\,d\tau_1\ldots d\tau_N, \tag{1·3}$$

where
$$f_{ij} = f(r_{ij}) = e^{-v(r_{ij})/kT} - 1; \tag{1·4}$$

or
$$Q_\tau = \iint\cdots\int[1 + \sum_{i>j} f_{ij} + \Sigma f_{ij} f_{i'j'} + \ldots]\,d\tau_1\ldots d\tau_N. \tag{1·5}$$

The usual procedure consists in evaluating the lowest terms of this series; the first sum corresponds to collisions of pairs of molecules, the second to triple encounters, etc. It is evident that this method is valid only for rarified gases and can never lead to a satisfactory theory of the condensed state.

---

* After finishing this paper we have received a manuscript of Professor Uhlenbeck and Dr B. Kahn, "On the theory of condensation", which will be published in *Physica*. We have to thank the authors for sending us this paper which is supplementary to our own work, chiefly concerning the extension of the theory to quantum statistics.

Mayer's intention is a formula holding for any density. He uses, therefore, another principle, first adopted by Ursell (1927), for ordering the terms in (1·5). To illustrate his idea we write down any product of the f_{ij} as it appears in the terms of (1·5), for instance

$$f_{12}f_{23}f_{45}f_{67}f_{78}f_{68}.$$

As only those parts of the phase space contribute to the integral for which the distances of all pairs of molecules which appear in the term are small we can describe the contribution of this term by saying: it consists of *one* cluster of two molecules (4, 5), *two* of three (1, 2, 3 and 6, 7, 8), whereas *all other* molecules are "free", or form clusters of one molecule.

To each cluster of l molecules there belongs an integral

$$b_l = \frac{1}{l!}\iint\ldots\int \Sigma\Pi f_{ij}\,d\tau_1\ldots d\tau_{l-1}, \tag{1·6}$$

where the factor $l!$ has been added for convenience and the integration is extended over the configuration volume of all but one of the l molecules, since evidently the integration over the last one yields simply the factor V (volume).

Each term of Q_τ consists of m_1 clusters of 1 molecule, m_2 clusters of 2, m_3 clusters of 3, etc., where

$$1m_1 + 2m_2 + 3m_3 + \ldots = \Sigma l m_l = N$$

is the total number of molecules. As the $l!$ permutations of the molecules in a cluster and the $m_l!$ permutations of the m_l equal clusters of these molecules correspond to the same arrangement, the weight of this is

$$\frac{N!}{(1!)^{m_1}(2!)^{m_2}\ldots m_1!\,m_2!\ldots} = N!\prod_l \frac{1}{(l!)^{m_l}m_l!}.$$

Therefore we get
$$\frac{Q_\tau}{N!} = \sum_{\substack{m_l\\ \Sigma l m_l = N}}\prod_l \frac{(Vb_l)^{m_l}}{m_l!} = \sum_{m_l}\prod_l \frac{(Nvb_l)^{m_l}}{m_l!}, \tag{1·7}$$

where
$$v = \frac{V}{N} \tag{1·8}$$

is the volume per molecule.

We shall consider first the integrals b_l and then evaluate the partition function Q_τ by means of a complex integration.

2. THEOREM ON THE REDUCTION OF CLUSTER INTEGRALS*

For a given cluster of size l there are a number of configurations, characterized by a product of the f_{ij}, such that this product cannot be split into two factors

$$F_{n_1 n_2 \ldots p}\, F_{p m_1 m_2 \ldots},$$

with *one* common index only. The sum of the corresponding integrals over the configuration space, divided by $(l-1)!$ (for convenience) is defined as the *irreducible integral* β_{l-1}.

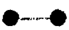
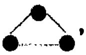
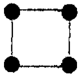
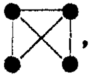
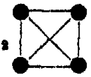
The simplest irreducible integrals are

$$\left.\begin{aligned} \beta_1 &= \int_0^\infty 4\pi r^2 f(r)\,d\tau, \\ \beta_2 &= \tfrac{1}{2}\iint f(r_{12})f(r_{23})f(r_{31})\,d\tau_1\,d\tau_2, \\ \beta_3 &= \tfrac{1}{6}(3\beta_{30}+6\beta_{31}+\beta_{32}), \end{aligned}\right\} \qquad (2\cdot 1)$$

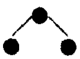
where

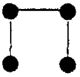
$$\left.\begin{aligned} \beta_{30} &= \iiint f_{12}f_{23}f_{34}f_{41}\,d\tau_1\,d\tau_2\,d\tau_3, \\ \beta_{31} &= \iiint f_{12}f_{13}f_{14}f_{23}f_{34}\,d\tau_1\,d\tau_2\,d\tau_3, \\ \beta_{32} &= \iiint f_{12}f_{13}f_{14}f_{23}f_{24}f_{34}\,d\tau_1\,d\tau_2\,d\tau_3. \end{aligned}\right\} \qquad (2\cdot 2)$$

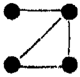
They correspond evidently to the following figures of "bonds":

β_1 , β_2 , β_{30} , β_{31} , β_{32} 

whereas figures as







belong to reducible integrals.

We shall say: an irreducible integral β_ν belongs to an "irreducible cluster" (of size $\nu+1$) with the index ν. It is useful to consider ν of the $\nu+1$ molecules of the cluster as numbered $(1, 2, \ldots, \nu)$ and one unnumbered; this last one is then used for "connecting" the irreducible cluster to another one by identifying it with a numbered particle of the other cluster. We define below such an aggregate of ν numbered particles of an irreducible cluster of size $\nu+1$ as an "incomplete cluster".

---

* This theorem has been found and proved by J. E. Mayer and Maria Goeppert-Mayer; but no complete proof has been published. We think that our proof is somewhat simpler than that which we tried to reconstruct from Mayer's indications.

Each cluster of the size l can be considered as composed of irreducible clusters of the indices $\nu = 0, 1, 2, \ldots, l-1$, where the irreducible cluster of index ν may appear μ_ν times. Therefore we can write

$$l!\,b_l = \sum_{\mu_\nu} p_l(\mu_1, \mu_2, \ldots) \prod_\nu (\nu!\,\beta_\nu)^{\mu_\nu}, \tag{2·3}$$

where the factors $l!$ and $\nu!$ are only a consequence of the definitions of the b_l and β_ν.

Each term of (2·3) is the integral of a product of the f_{ij} of the form

$$F_{l_1 l_2 \ldots p}\, F_{m_1 m_2 \ldots q}\, F_{pq n_1 n_1 \ldots r}\, F_{r\sigma_1 \sigma_2 \ldots} \cdots F_{y s_1 s_2 \ldots z}, \tag{2·4}$$

multiplied with the number of possibilities $p_l(\mu_1, \mu_2, \ldots)$ of interchanging the particles without destroying the form of the integrand.

As integration over the ν_1 particles $l_1 l_2 \ldots$ makes the corresponding integral factor of (2·4) independent of p, we count p not to the first incomplete cluster, but in the example (2·4) to the third; in the same way we count the particle q to the third, r to the fourth, etc. Then in the last factor the last particle z plays a special role. We define the "incomplete clusters" which compose the given one, by the groups $(l_1 l_2, \ldots)$ of number ν_1, $(m_1 m_2 \ldots)$ of number ν_2 and so on, but for the last $(y s_1 s_2 \ldots z)$ of number $\nu_a + 1$. Then the number of possibilities of distributing the l particles over the incomplete clusters of $\nu_1, \nu_2, \ldots, \nu_a + 1$ particles is given by

$$\frac{l!}{\prod_\nu (\nu!)^{\mu_\nu} \prod_\nu (\mu_\nu)!\, \nu_a!}\,\frac{1}{\nu_a+1}, \tag{2·5}$$

where we have split $(\nu_a+1)!$ into the factors $\nu_a!(\nu_a+1)$. We remark here that any arbitrary cluster can be taken as the "last" one (with index ν_a).

$p_l(\mu_1, \mu_2, \ldots)$ is therefore the factor (2·5) multiplied by the number κ of possibilities of connecting the incomplete clusters. We define $q(\mu_1, \mu_2, \ldots)$ as this number κ divided (for convenience) by $\nu_a + 1$. Then (2·3) can be written

$$b_l = \sum_{\mu_\nu} q(\mu_1, \mu_2, \ldots) \prod_{\substack{\nu \\ \Sigma \nu\mu_\nu = l-1 \\ \nu}} \frac{(\beta_\nu)^{\mu_\nu}}{(\mu_\nu)!}, \quad q(\mu_1, \mu_2, \ldots) = \frac{\kappa}{\nu_a+1}. \tag{2·6}$$

In forming the factor (2·5) we have taken into account that two irreducible clusters of the same index can be exchanged, i.e. we have counted as identical such distributions of the particles over the incomplete clusters which are obtained by exchanging all particles of two clusters of equal index. We must therefore now consider two possibilities of connecting the incomplete clusters, which are obtained by exchanging two clusters of the same index, as different. For this reason it is more convenient to introduce a different notation. We numerate the irreducible clusters in any order from 1 to m, where m is the total number of irreducible clusters, and characterize any configuration by the indices of the irreducible clusters $\nu_1, \nu_2, \ldots, \nu_m$, where several ν_i may be equal. Then

$$\sum_{i=1}^{m} \nu_i = l-1, \tag{2·7}$$

and q is now to be considered as a function of the ν_i:

$$q = q(\nu_1, \nu_2, \ldots, \nu_m). \tag{2·8}$$

We assume now that the l particles are distributed over the incomplete clusters in a definite way. Each irreducible cluster—say of index ν_i—contains ν_i numbered particles (those of the incomplete cluster) and one unnumbered particle. (In the example (2·4) p is the unnumbered particle of the first cluster, q that of the second, and so on; only the ν_a+1 particles of the "last" cluster are all numbered.) We "attach" one irreducible cluster ν_i to another ν_j by identifying the unnumbered particle of the irreducible cluster ν_i with any numbered particle of the irreducible cluster ν_j.

We consider first the simplest case of forming such attachments between the irreducible clusters in a simple chain. We start with the "last" cluster containing ν_a+1 numbered particles and attach another cluster ν_b, which can be done in ν_a+1 different ways, since we can identify the unnumbered particle of the cluster ν_b with any one of the ν_a+1 particles of the last cluster. Correspondingly we obtain a factor ν_b, when we attach a third cluster ν_c to the second and so on (compare fig. 1a). The total number of possibilities of forming such a chain in any given order of the irreducible clusters is $(\nu_a+1)\prod_i \nu_i$, $(i \neq a,\ i \neq x)$, where ν_x is the cluster attached last. We obtain the same product $(\nu_a+1)\prod_i \nu_i$, $(i \neq a,\ i \neq x)$, if we exchange the $m-2$ clusters in the chain, keeping the clusters ν_a and ν_x fixed. Dividing by ν_a+1 we obtain as contribution to q

$$(m-2)!\prod_{\substack{i \neq a \\ i \neq x}} \nu_i. \tag{2·9}$$

We take always the cluster ν_a as the starting point for the process of attaching; but it is not necessary to attach just one cluster to ν_a, we may attach two or more. If two clusters are attached to ν_a—say ν_b and ν_c—one obtains a chain in which the cluster ν_a is not at the end, as shown in fig. 1*b*.

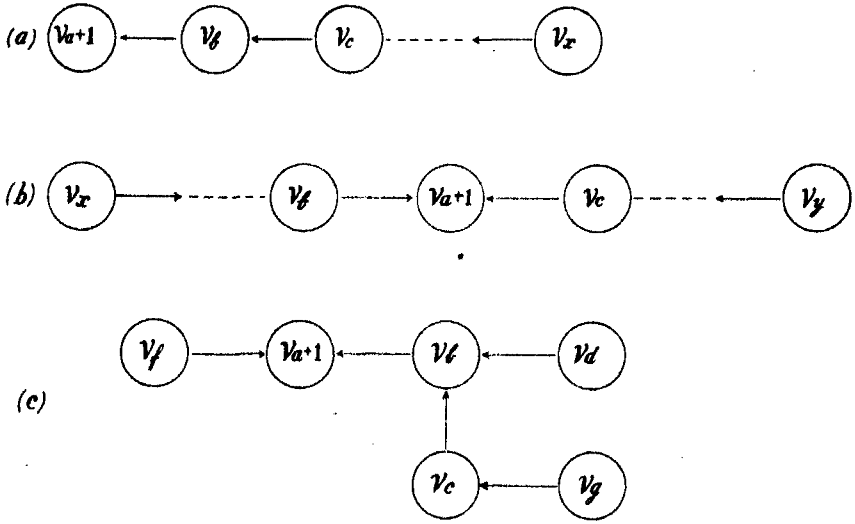


FIG. 1

In general we are not confined to attaching each cluster to the immediately previously attached cluster, but we can attach them to any one of the previously attached clusters, starting a "side-chain". In this case the factor ν_i may appear several times, though the total number of factors is always equal to $m-2$, since after $m-1$ steps all clusters are attached and a factor ν_a+1 always cancels. (In the case illustrated in fig. 1*c*, for example, we obtain $(\nu_a+1)\,\nu_b^2\nu_c$.) It is to be noted that ν_a+1 appears always instead of ν_a; the following formulae should be understood in this sense. Therefore, q can be written as the sum of such products

$$q(\nu_1, \nu_2, \dots \nu_m) = \sum_{\substack{n_i \\ \Sigma n_i = m-2}} f(m; n_1, n_2, \dots) \prod_i \nu_i^{\,n_i}. \qquad (2{\cdot}10)$$

$f(m; n_1, n_2 \dots)$ is here the number of possible arrangements in which n_1 clusters are attached to the cluster ν_1, n_2 to the cluster ν_2, and so on, and n_a+1 to the cluster ν_a+1 (since one factor ν_a+1 is cancelled, cf. (2·6)).

We shall prove by induction that

$$f(m; n_1, n_2, \ldots) = \frac{(m-2)!}{\prod_{i=1} n_i!}. \tag{2·11}$$

For $m = 2$ or 3 this formula has already been proved; for in this case it reduces to the formula (2·9) for a simple chain. We, therefore, assume (2·11) is true for all values of m smaller than m_0 and prove that then it also holds for m_0.

We first consider $m_0 - 1$ clusters leaving aside one of the clusters at the end of the chain or of a side-chain (i.e. a cluster the index ν_x of which does not appear in the product). Since there are at least two such clusters, we can avoid leaving aside the cluster $\nu_a + 1$. We now connect these $m_0 - 1$ clusters in such a way that n_i factors ν_i appear; however, since one cluster is missing there will be one factor less—say the factor ν_j, which appears $n_j - 1$ times instead of n_j times (where $n_j \geqslant 1$). The number of possibilities of doing so is according to (2·11)

$$n_j \frac{(m_0 - 3!)}{\prod_i n_i!},$$

where we have written $n_j/n_j!$ for $1/(n_j - 1)!$; in this form it also holds if $n_j = 0$. The last cluster is now to be attached to the cluster ν_j supplying the missing factor ν_j. Summing over all j we obtain

$$f(m_0; n_1, n_2, \ldots) = \sum_j n_j \frac{(m_0 - 3)!}{\prod_i n_i!}.$$

Since $\sum_j n_j = m_0 - 2$ (compare (2·10)) we obtain formula (2·11), which thus is proved.

Inserting (2·11) in (2·10) we obtain

$$q(\nu_1, \nu_2, \ldots, \nu_m) = \sum_{\substack{n_i \\ \Sigma n_i = m-2}} (m-2)! \prod_i \frac{\nu_i^{n_i}}{n_i!} = \left(\sum_i \nu_i\right)^{m-2}.$$

Remembering that in each formula ν_a must be replaced by $\nu_a + 1$ this expression is according to (2·7)

$$q(\nu_1, \nu_2, \ldots, \nu_m) = l^{m-2}.$$

Inserting this expression in (2·6) and observing that $\Sigma \mu_\nu = m$, we obtain finally the

THEOREM: *Each cluster integral can be split into a sum of products of irreducible integrals corresponding to the formula*

$$b_l = l^{-2} \sum_{\substack{\mu_\nu \\ \Sigma \nu \mu_\nu = l-1}} \prod_\nu \frac{(l\beta_\nu)^{\mu_\nu}}{\mu_\nu!}. \tag{2·12}$$

3. INTEGRAL REPRESENTATION AND ASYMPTOTIC EXPRESSIONS

The formula (1·7) for Q_τ and formula (2·12) for b_l are of the same type:

$$F(M, I, x_1, x_2, \ldots) = \sum_{\substack{m_l \\ \Sigma l m_l = M}} \prod_l \frac{(I x_l)^{m_l}}{m_l!} = F(M, I, x). \tag{3·1}$$

With this notation these two formulae can be written

$$\left.\begin{aligned} \frac{Q_\tau}{N!} &= F(N, N, vb_1, vb_2, \ldots) = F(N, N, vb), \\ l^2 b_l &= F(l-1, l, \beta_1, \beta_2, \ldots) = F(l-1, l, \beta). \end{aligned}\right\} \tag{3·2}$$

These expressions can be represented by complex integrals for which asymptotic expressions for large I can be derived.

The functions $F(M, I, x)$ are the coefficients in the expansion of

$$\exp(I \Sigma x_\nu \xi^\nu)$$

in a series of powers of ξ. For, expanding the exponential function we obtain

$$\begin{aligned} e^{I \Sigma x_\nu \xi^\nu} &= \sum_m \frac{1}{m!} \left(\sum_\nu I x_\nu \xi^\nu\right)^m \\ &= \sum_m \frac{1}{m!} \sum_{\substack{\mu_\nu \\ \Sigma \mu_\nu = m}} m! \prod_\nu \frac{(I x_\nu \xi^\nu)^{\mu_\nu}}{\mu_\nu!}. \end{aligned}$$

Here the second sum is to be taken over all sets of values of the μ_ν for which $\Sigma \mu_\nu = m$. Since the factor $m!$ cancels m appears only in this restriction for the values of the μ_ν, and summing over all m, the restriction is removed. We can therefore write

$$e^{I \Sigma x_\nu \xi^\nu} = \sum_{\mu_\nu} \prod_\nu \frac{(I x_\nu)^{\mu_\nu} \xi^{\nu \mu_\nu}}{\mu_\nu!} = \sum_{\mu_\nu} \xi^{\sum_\nu \nu \mu_\nu} \prod_\nu \frac{(I x_\nu)^{\mu_\nu}}{\mu_\nu!}.$$

We now take together all sets of values μ_ν, for which $\sum_\nu \nu \mu_\nu = M$; we can then take $\xi^{\Sigma \nu \mu_\nu} = \xi^M$ in front of the sum over μ_ν, and obtain

$$e^{I \Sigma x_\nu \xi^\nu} = \sum_M \xi^M \sum_{\substack{\mu_\nu \\ \Sigma \nu \mu_\nu = M}} \prod_\nu \frac{(I x_\nu)^{\mu_\nu}}{\mu_\nu!}.$$

Using the definitions

$$G_\lambda(\xi; x_1, x_2, \ldots) = G_\lambda(\xi, x) = \sum_{\nu=1}^{\infty} x_\nu \xi^\nu \nu^\lambda \tag{3·3}$$

and (3·1) we may write

$$e^{I\Sigma x_\nu \xi^\nu} = e^{IG_0(\xi, x)} = \sum_M F(M, I, x)\,\xi^M. \tag{3·4}$$

Applying Cauchy's integral theorem, F can be expressed as the integral

$$F(M, I, x) = \frac{1}{2\pi i}\oint \frac{e^{IG_0(\xi, x)}}{\xi^{M+1}}\,d\xi. \tag{3·5}$$

We thus obtain the *integral representation for the* $F(M, I, x)$

$$F(M, I, x) = \frac{1}{2\pi i}\oint e^{IG_0(\xi, x)-(M+1)\ln \xi}\,d\xi, \tag{3·6}$$

where the integration is to be extended over a path inside the region of regularity of $\exp(IG_0(\xi, x))$ enclosing the point $\xi = 0$. It is obvious that the exponential function is regular inside the radius of convergence of $G_0(\xi, x)$ with the exception of the point $\xi = 0$.

We shall now derive the asymptotic expressions for the integrals (3·6) by the *method of steepest descent*. We shall choose as path of integration a circle around the origin and therefore write

$$\xi = Ze^{i\phi}, \quad d\xi = iZe^{i\phi},$$

$$\ln \xi = i\phi + \ln Z.$$

Substituting these expressions in (3·6) the integral can be written in the form

$$\left.\begin{aligned} F(M, I, x) &= \frac{1}{2\pi}\int_{-\pi}^{+\pi} e^{IG_0(Ze^{i\phi}, x)-(M+1)\,i\phi-(M+1)\ln Z}\, Ze^{i\phi}\,d\phi, \\ &= \frac{1}{2\pi}\int_{-\pi}^{+\pi} e^{f(Ze^{i\phi})}\,d\phi, \\ f(Ze^{i\phi}) &= IG_0(Ze^{i\phi}, x) - M\ln(Ze^{i\phi}). \end{aligned}\right\} \tag{3·7}$$

We now choose the radius of the circle of integration Z such that the function $f(Ze^{i\phi})$ of ϕ in the exponent has a maximum at the real axis (for $\phi = 0$). Since according to (3·3)

$$\frac{\partial}{\partial\phi}[G_\lambda(Ze^{i\phi}, x)] = iG_{\lambda+1}(Ze^{i\phi}, x), \tag{3·8}$$

we obtain as equation for Z

$$\left[\frac{\partial}{\partial\phi} f(Ze^{i\phi})\right]_{\phi=0} = i[IG_1(Z, x) - M] = 0,$$

or

$$G_1(Z, x) = \frac{M}{I}. \tag{3·9}$$

The point Z on the real axis is a saddle point of the function $f(\xi)$; $f(\xi)$ has a maximum perpendicular to the real axis and a minimum parallel to it.

The higher derivatives of $f(\xi)$ at the saddle point with regard to ϕ can easily be obtained, taking into account the relation (3·8); one finds

$$\left[\frac{\partial^\nu}{\partial\phi^\nu}f(Ze^{i\phi})\right]_{\phi=0} = Ii^\nu G_\nu(Z, x),$$

and we can write (3·7)

$$F(M, I, x) = \frac{1}{2\pi}e^{f(Z)}\int_{-\pi}^{+\pi}\exp\left[-IG_2(Z,x)\,\phi^2/2 + I\sum_{\nu=3}^{\infty} i^\nu G_\nu(Z,x)\,\phi^\nu/\nu!\right]d\phi.$$

Substituting $\phi^2 = y^2 IG_2/2$ we obtain

$$F(M, I, x) = \frac{1}{2\pi}e^{f(Z)}\sqrt{\frac{2}{IG_2(Z,x)}} \times \int_{-\pi\sqrt{\frac{IG_2}{2}}}^{+\pi\sqrt{\frac{IG_2}{2}}}\exp\left[-y^2 + I\sum_{\nu=3}^{\infty} i^\nu G_\nu(Z,x)\,y^\nu/\nu!(IG_2/2)^{\nu/2}\right]dy. \tag{3·10}$$

If I is large all terms in the exponential except y^2 are negligible and the integral can be extended from $-\infty$ to $+\infty$ with negligible error. The integral is the well-known error function and yields a factor $\sqrt{\pi}$;* substituting according to (3·7) for $f(Z)$ we obtain the *asymptotic expression*

$$F(M, I, x) = \frac{e^{IG_0(Z,x)}}{Z^M\sqrt{\{2\pi IG_2(Z, x)\}}}, \quad I \gg 1. \tag{3·11}$$

In neglecting all terms in the exponent of (3·10) except y^2 we have assumed tacitly that the G_ν are of smaller order than I. This need not be the case; in general we must add the condition

$$\frac{1}{\nu!}\frac{IG_\nu(Z, x)}{[IG_2(Z, x)/2]^{\nu/2}} \ll 1, \quad \nu \geq 3. \tag{3·12}$$

---

* We have neglected here the possibility that other maxima might occur apart from the maximum at $\phi = 0$. However, if all x_ν are positive it can be shown quite easily that these maxima will contribute negligible amounts. For the exponent of the integrand in (3·10) can be written:

$$f(Ze^{i\phi}) - f(Z) = I\Sigma x_\nu Z^\nu[(\cos\nu\phi - 1) + i\sin\nu\phi] - iM\phi.$$

The real part of this expression is negative and of the order I, except for $\phi = 0$.

Specializing for the two cases (3·2) we obtain from (3·11)

$$\frac{Q_\tau}{N!} = F(N, N, vb) = \frac{e^{NG_0(Z, vb)}}{Z^N\sqrt{[2\pi N G_2(Z, vb)]}}, \quad N \gg 1 \tag{3·13}$$

$$l^2 b_l = F(l-1, l, \beta) = \frac{e^{lG_0(\rho, \beta)}}{\rho^{l-1}\sqrt{\{2\pi l G_2(\rho, \beta)\}}}, \quad l \gg 1 \tag{3·14}$$

$$G_\lambda(Z, vb) = v \sum_{\nu=1}^{\infty} \nu^\lambda b_\nu Z^\nu, \tag{3·15}$$

$$G_\lambda(\rho, \beta) = \sum_{\nu=1}^{\infty} \nu^\lambda \beta_\nu \rho^\nu, \tag{3·16}$$

where Z and ρ are defined by the equations:

$$G_1(Z, vb) = 1, \tag{3·17}$$

$$G_1(\rho, \beta) = 1. \tag{3·18}$$

β_ν has the same dimension as $b_\nu v$, namely that of v^ν; it follows from (3·17), (3·18) that Z and ρ have the dimension of v^{-1} (or density).

We shall use the notation $$v_s = \frac{1}{\rho}. \tag{3·19}$$

This volume, which depends on the temperature but not on the pressure, will play an important role in the theory of condensation.

The quantity Z, as given by (3·17), has also a simple physical meaning, which can be given only after deriving the thermodynamical properties of the system.

4. THERMODYNAMICAL PROPERTIES OF THE GAS

The equation (3·13) enables us to express the thermodynamic potentials in terms of the series $G_\lambda(Z, vb)$. Neglecting all terms but those of the highest order in N one gets

$$\frac{1}{N}\ln\frac{Q_\tau}{N!} = G_0(Z, vb) - \ln Z, \quad G_1(Z, vb) = 1. \tag{4·1}$$

The thermodynamical properties of the systems can be determined from the free energy A (as introduced by Helmholtz) which is connected with the partition function by the relation

$$A = -kT\ln Q. \tag{4·2}$$

If we introduce here the values (1·1) and (4·1) we get with $R = kN$

$$A = RT\left\{\ln\left(\frac{h^2}{2\pi mkT}\right)^{\frac{3}{2}} - G_0(Z) + \ln Z\right\}. \tag{4·3}$$

The pressure is $$P = -\left(\frac{\partial A}{\partial V}\right)_T = \frac{RT}{V} v \frac{\partial}{\partial v}\{G_0(Z) - \ln Z\}. \tag{4·4}$$

In the condensed state the b_ν may depend on v; but in the gaseous state we can assume them being independent of v. Then the differentiation in (4·4) can be easily performed; from (3·15) one has

$$Z\frac{d}{dZ}G_\lambda(Z) = G_{\lambda+1}(Z);$$

therefore for $\lambda = 0$ $$Z\frac{dG_0(Z)}{dZ} = G_1(Z) = 1. \tag{4·5}$$

Now we get

$$\begin{aligned} v\frac{\partial}{\partial v}\{G_0(Z) - \ln Z\} &= v\frac{\partial}{\partial v}(v \sum_\nu b_\nu Z^\nu - \ln Z) \\ &= v \sum_\nu b_\nu Z^\nu + \frac{v}{Z}\left(Z\frac{\partial G_0(Z)}{dZ} - 1\right)\frac{dv}{\partial Z}. \end{aligned}$$

The last bracket vanishes according to (4·5), and one gets

$$v\frac{\partial}{\partial v}\{G_0(Z) - \ln Z\} = G_0(Z). \tag{4·6}$$

Introducing this in (4·4) one finds the *equation of state*

$$PV = RTG_0(Z). \tag{4·7}$$

Gibbs' thermodynamic potential is

$$F = A + PV = RT\left\{\ln\left(\frac{h^2}{2\pi mkT}\right)^{\frac{3}{2}} + \ln Z\right\}. \tag{4·8}$$

From (4·7) one gets with (4·5)

$$V^2\frac{dP}{dV} = RTG_1(Z)\frac{d\ln Z}{d\ln v} = RT\frac{d\ln Z}{d\ln v}. \tag{4·9}$$

The problem of finding the virial coefficients will be solved, if we can express $G_0(Z)$ in terms of the volume. This will be done in the next paragraph.

5. DETERMINATION OF THE VIRIAL COEFFICIENTS

In order to find an expression of $G_0(Z)$ in terms of the volume we consider the following functions

$$H_\lambda^m(z, x_1, x_2, \ldots) = H_\lambda^m(z, x) = \sum_{N=1}^{\infty} \frac{F(N-m, N, x)}{N^\lambda} z^N \tag{5·1}$$

where F is defined by (3·1); we prove the following

THEOREM

$$H_0^m(z, x) = \frac{\xi_0^m}{1 - G_1(\xi_0, x)} - \delta(m, 0) \tag{5·2}$$

$$\delta(m, 0) = \begin{cases} 1 \text{ for } m = 0 \\ 0 \text{ for } m \neq 0, \end{cases}$$

where ξ_0 is defined by $$z = \xi_0 e^{-G_0(\xi_0, x)} \tag{5·3}$$

and $$G_\lambda(\xi_0, x) = \sum_{\nu=1}^{\infty} \nu^\lambda x_\nu \xi_0^\nu, \tag{5·4}$$

so that $$\xi_0 \frac{d}{d\xi_0} G_\lambda(\xi_0, x) = G_{\lambda+1}(\xi_0, x). \tag{5·5}$$

We shall prove the theorem for small values of z; it then follows that it is valid in general.

Inserting the complex integral (3·5) for the F-functions in (5·1) we obtain

$$H_0^m(z, x) = \frac{1}{2\pi i} \sum_{N=1}^{\infty} z^N \oint \frac{e^{NG_0(\xi, x)}}{\xi^{N-m+1}} d\xi, \tag{5·6}$$

where the path of integration must include the point $\xi = 0$. We choose the path in such a way that

$$\left| z \frac{e^{G_0(\xi)}}{\xi} \right| < 1. \tag{5·7}$$

This is possible for small values of z as shown below. We can now exchange integration and summation in (5·6) and obtain

$$H_0^m(z, x) = \frac{1}{2\pi i} \oint \frac{z e^{G_0(\xi)}}{\xi - z e^{G_0(\xi)}} \xi^{m-1} d\xi, \tag{5·8}$$

This integral has a singularity for a value ξ_0 where (5·3) is satisfied. The residuum of the singularity is obtained by inserting the derivative of the denominator in (5·8) in place of the denominator; with (5·5) and (5·3) we obtain for the residuum

$$\frac{z e^{G_0(\xi_0)}}{1 - z e^{G_0(\xi_0)} G_1(\xi_0)/\xi_0}\, \xi_0^{m-1} = \frac{\xi_0^m}{1 - G_1(\xi_0)}.$$

If $m = 0$, there is another singularity at $\xi = 0$ with the residuum -1. Thus equation (5·2) is proved, provided there is one and only one point ξ_0 inside the path of integration for which (5·3) is satisfied.

Now, the function
$$e^{G_0(\xi)}/\xi \tag{5·9}$$

has a pole of the first order at $\xi = 0$. This pole is surrounded by curves along which the absolute value of (5·9) is constant. Provided the absolute value is sufficiently high these curves are closed around $\xi = 0$; therefore, if $|z|$ is sufficiently small we can choose the path of integration according to (5·7) just outside the curve along which

$$\left|\frac{e^{G_0(\xi)}}{\xi}\right| = \frac{1}{|z|},$$

in such a way that there is only one such curve inside the path of integration. On this curve there is one and only one point for which (5·3) is satisfied.

From (5·3) we obtain with (5·5)

$$\frac{dz}{z} = \frac{d\xi_0}{\xi_0}[1 - G_1(\xi_0)]. \tag{5·10}$$

Multiplying (5·2) with (5·10) and integrating we find

$$\int_0^z H_0^m(z', x)\frac{dz'}{z'} = \int_0^{\xi_0}\left[\xi^{m-1} - \delta(m, 0)\frac{1 - G_1(\xi)}{\xi}\right]d\xi.$$

The left-hand side yields the series H_1^m according to (5·1). The right-hand side gives for $m > 0$

$$H_1^m(z, x) = \frac{\xi_0^m}{m}, \quad m > 0. \tag{5·11}$$

For $m = 0$ the terms $1/\xi$ cancel and we obtain with (5·5)

$$H_1^0(z, x) = G_0(\xi_0, x). \tag{5·12}$$

Repeating the integration in the same way, we obtain from (5·11)

$$H_2^m(z, x) = \frac{\xi_0^m}{m}\left[\frac{1}{m} - G_{0,m}(\xi_0, x)\right], \quad m > 0 \tag{5·13}$$

$$G_{0,m}(\xi_0, x) = \sum_{\nu=1}^{\infty} \frac{\nu}{\nu + m} x_\nu \xi_0^\nu.$$

According to (3·2), (3·3) the series $G_\lambda(z, vb)$ can be represented by H-functions:

$$G_\lambda(Z, vb) = vH^1_{2-\lambda}(Z, \beta). \tag{5·14}$$

For $\lambda = 0, 1$ we obtain with (5·11), (5·13) (with $\xi_0 = y$)

$$G_1(Z, vb) = vH^1_1(Z, \beta) = yv \tag{5·15}$$

$$G_0(Z, vb) = vH^1_2(Z, \beta) = yv[1 - G_{0,1}(y, \beta)] = yv\left[1 - \sum_{\nu=1}^{\infty} \frac{\nu}{\nu+1}\beta_\nu y^\nu\right] \tag{5·16}$$

where y is given by (5·3) $\qquad Z = ye^{-G_0(y, \beta)}$. (5·17)

The root of the equation $G_1(Z, vb) = 1$ is, according to (5·15), given by (5·17) with

$$y = \frac{1}{v}. \tag{5·18}$$

The *equation of state* can now be obtained by inserting (5·16) in (4·7)

$$PV = RTG_0(Z, vb) = RT\left[1 - \sum_{\nu=1}^{\infty} \frac{\nu}{\nu+1}\beta_\nu \frac{1}{v^\nu}\right]. \tag{5·19}$$

Thus the virial coefficients are given by the irreducible integrals β_ν multiplied with $-\nu/\nu+1$.*

6. CONDENSATION

The formulae of the last section can also be applied to the series

$$H^0_\lambda(r, vb) = \sum_{N=1}^{\infty} \frac{Q_{\tau, N}}{N!}\frac{r^N}{N^\lambda}. \tag{6·1}$$

For according to (3·2) these are again H-functions. We obtain from (5·2), (5·12) and (5·3) (replacing z by r and ξ_0 by z):

$$\sum_{N=1}^{\infty} \frac{Q_{\tau, N}}{N!} r^N = H^0_0(r, vb) = \frac{1}{1 - G_1(z, vb)} - 1, \tag{6·2}$$

$$\sum_{N=1} \frac{Q_{\tau, N}}{N!}\frac{r^N}{N} = H^0_1(r, vb) = G_0(z, vb), \tag{6·3}$$

$$r = ze^{-G_0(z, vb)}; \quad \frac{dr}{r} = \frac{dz}{z}[1 - G_1(z, vb)], \tag{6·4}$$

---

* This theorem has been proved independently by Mayer and Harrison, Uhlenbeck and Kahn, and ourselves. The proof given here (which is, in fact, a special case of Lagrange's theorem) is essentially that of Uhlenbeck and Kahn; we are very grateful for being permitted to use this method which is simpler than the others.

whereas the corresponding expressions for the $G_\lambda(z, vb)$ are, with (5·2), (5·15–5·17):

$$G_0(z, vb) = vH^1_2(z, \beta) = yv[1 - G_{0,1}(y, \beta)] = yv\left[1 - \sum_{\nu=1}^{\infty} \frac{\nu}{\nu+1} \beta_\nu y^\nu\right]. \tag{6·5}$$

$$G_1(z, vb) = vH^1_1(z, \beta) = yv, \tag{6·6}$$

$$G_2(z, vb) = vH^1_0(z, \beta) = \frac{yv}{1 - G_1(y, \beta)}, \tag{6·7}$$

$$z = ye^{-G_0(y, \beta)}; \quad \frac{dz}{z} = \frac{dy}{y}[1 - G_1(y, \beta)]. \tag{6·8}$$

We now use the fact that by definition (1·2) the $Q_{\tau, N}$ are all positive. Professor G. E. Uhlenbeck has kindly communicated to us an idea due to Dr Kahn: The radius of convergence R of the power series in r (6·2) and (6·3) is determined by the theorem of Cauchy-Hadamard as

$$\frac{1}{R} = \lim_{N \to \infty} \left(\frac{Q_{\tau, N}}{N!}\right)^{1/N}.$$

This theorem supplies just the proper tool for handling the partition function. For, taking logarithms we find exactly the quantity which we need in order to obtain an expression for the free energy

$$\lim_{N \to \infty} \frac{1}{N} \ln \frac{Q_{\tau, N}}{N!} = -\ln R. \tag{6·9}$$

Let Z be the value corresponding to R; we have with (6·4)

$$\lim_{N \to \infty} \frac{1}{N} \ln \frac{Q_{\tau, N}}{N!} = G_0(Z, vb) - \ln Z. \tag{6·10}$$

This equation is formally identical with (4·1). However Z is here defined in a different way.

Let us now consider a region in the z-plane such that

(α) The $G_\lambda(z, vb)$ are regular;

(β) $G_1(z, vb) \neq 1$.

Then (6·4) defines a conformal correspondence of the z-plane and the r-plane, and the series $H^0_0(r, vb)$ is regular in the corresponding region of the r-plane. If H^0_0 is singular, one of the conditions (α), (β) must be violated.

The reverse also holds; if (β) is violated, it is obvious from (6·2) that H^0_0

has a pole. In order to show that H_0^0 is singular if (α) is violated, we consider the two equations (6·3), (6·4) which we may write

$$G_0(z, vb) = H_1^0(r, vb)$$

$$z = re^{H_1^0(r, vb)}; \quad \frac{dz}{z} = \frac{dr}{r}[1+H_0^0(r, vb)].$$

In the same way as above we find that if $G_0(z, vb)$ is singular, either $H_0^0(r, vb)$ is singular, or

$$H_0^0(r, vb) = -1. \tag{6·11}$$

Since all coefficients of $H_0^0(r, vb)$ are positive (compare (1·2)) (6·11) cannot be satisfied on the positive r-axis. For the same reason the singularity which determines the radius of convergence R is on the positive axis. We, therefore, can choose a region inside the radius of convergence R, including the positive axis from $r = 0$ to $r = R$, in which there is no point for which (6·11) is satisfied.

We can now conclude that this region is conformally represented by a region in the z-plane, in which the $G_\lambda(z, vb)$ are regular. In particular the positive r-axis is represented by the positive z-axis and to the point R there corresponds a point Z which we can now define as the smallest positive z for which one of the conditions (α) and (β) is violated.

We now consider the equations (6·5)–(6·8); in the same way as above it follows that if $G_0(z, vb)$ has a singularity, one of the following conditions must be violated:

(a) The $G_\lambda(y, \beta)$ are regular;

(b) $G_1(y, \beta) \neq 1$.

The reverse argument is slightly different; we obtain from (6·6) and (6·7)

$$\left.\begin{aligned} G_1(y, \beta) &= 1 - \frac{G_1(z, vb)}{G_2(z, vb)}, \\ y = \frac{1}{v} G_1(z, vb); &\quad dy = v\frac{dz}{z} G_2(z, vb). \end{aligned}\right\} \tag{6·12}$$

If $G_1(y, \beta)$ is singular, either $G_1(z, vb)$ is singular or $G_2(z, vb) = 0$. If $G_2(z, vb) = 0$, it follows from (6·12) that $G_1(y, \beta) = -\infty$. In that case it follows from the equation (6·21) (which we shall prove later) that dP/dV tends to $-\infty$. We should have to deal with an incompressible body. We can exclude this case for physical reasons.

We can now argue as before that to the point $\bar{Z}$ for which $G_\lambda(z, vb)$ is singular, there corresponds a point $\overline{Y}$ which is the smallest positive y for which either condition (*a*) or condition (*b*) is violated.

Since the first coefficient of $G_2(z, vb)$ is positive ($b_1 = 1$) it follows that on the positive z-axis

$$G_2(z, vb) > 0 \quad \text{for } |Z| = Z. \tag{6·13}$$

The equations for the gaseous phase

We shall first consider the case (β). Then Z is given by

$$G_1(Z, vb) = v\Sigma\nu b_\nu Z^\nu = 1. \tag{6·14}$$

Since $b_1 = 1$, there will always be a root of (6·14) inside the region of regularity for sufficiently large values of v. From (6·6) we conclude that the corresponding point Y is given by

$$Y = \frac{1}{v}. \tag{6·15}$$

With (6·14) the expression (6·10) for the partition function is identical with (4·1). We obtain, therefore, all the equations of the gaseous state derived in the preceding sections, which in this way are proved to be valid in general since we have made no assumptions about the b_l or β_ν, provided we use the analytical continuations if the series in question diverge. (If some of the b_l or β_ν are negative it is of course possible that the series diverge, but that the functions represented remain regular on the positive axis.)

Condensation

Let $\overline{Z}$ be the smallest positive singular point of $G_1(z, vb)$, which tends to a finite limit as z approaches $\overline{Z}$

$$\lim_{z\to\overline{Z}} G_1(z, vb) = \frac{v}{v_s}. \tag{6·16}$$

Since $G_1(z, vb)$ increases monotonically according to (6·13) it follows that for $v > v_s$ there exists a root of (6·14) and we are in the region of the gaseous phase. For $v < v_s$, however, there is no root of (6·14) inside the region of regularity and Z is given by $\overline{Z}$.

$$Z = \overline{Z}, \quad v < v_s. \tag{6·17}$$

With (6·10), (4·2) and (4·4), we find

$$PV = RTG_0(\overline{Z}, vb) = RTv\Sigma b_\nu \overline{Z}^\nu, \quad v < v_s. \tag{6·18}$$

Since $\bar{Z}$ does not depend on the volume it follows that P remains constant. v_s therefore is the condensation volume. v_s can be obtained from (6·15)

$$\bar{Y} = \frac{1}{v_s} \tag{6·19}$$

where $\bar{Y}$ is the point corresponding to $\bar{Z}$.

We now have to distinguish between two cases, corresponding to the two conditions (a) and (b). If the condition (b) is violated $\bar{Y}$ is the smallest positive root of the equation

$$G_1(\bar{Y}, \beta) = 1. \tag{6·20}$$

In this case $\bar{Y}$ is identical with the quantity ρ defined by (3·18). If no such root exists inside the region of regularity of $G_1(y, \beta)$, $\bar{Y}$ is the smallest singular positive point of $G_1(y, \beta)$.

In the two cases the gas shows a distinctly different behaviour as we approach the condensation volume. From (5·19) it follows that for the gaseous phase

$$V^2 \frac{dP}{dV} = RT[G_1(Y, \beta) - 1], \quad Y = \frac{1}{v}, \quad v > v_s. \tag{6·21}$$

If $\bar{Y}$ is defined as the root of (6·20) it follows from that definition that dP/dV tends continuously towards zero as v approaches v_s.

If, however, $\bar{Y}$ is a singular point, so that $G_1(\bar{Y}, \beta) < 1$, dP/dV is discontinuous for $v = v_s$. In order to obtain agreement with the observed facts, we must assume that $\bar{Y}$ is a singular point.

If the singularities $\bar{Y}$ and $\bar{Z}$ are situated on the circle of convergence of the series $G_1(z, vb)$ and $G_1(y, \beta)$, the theory of condensation as presented here is in essence identical with Mayer's theory; we believe the above considerations show without doubt that his theory is correct in this case.

If, however, the singularities $\bar{Y}$ and $\bar{Z}$ are beyond the radii of convergence, the equations for the gas hold beyond these radii and we have to introduce the analytical continuations of these series.

The physical meaning of the quantities Y, Z and R can now easily be seen. That of Y is clear from equations (6·15) and (6·19). The physical meaning of Z can be seen from (4·8); if we divide (4·8) by the number of molecules N and write $R = Nk$, we get

$$Z = \left(\frac{2\pi mkT}{h^2}\right)^{\frac{3}{2}} e^{\mu/kT}, \tag{6·22}$$

$$\mu = \frac{F}{N}.$$

Thus Z is connected with the thermodynamic potential per molecule, μ, which is defined as the derivative with regard to N of the total energy E as function of N, V, S

$$dE = \mu dN + TdS - PdV. \tag{6·23}$$

The function E is found from (4·8)

$$E = A + TS = F - PV + TS = \mu N - PV + TS. \tag{6·24}$$

The last expression can be interpreted as Euler's condition for E being homogeneous in N, V, S.

Following a remark of Professor R. H. Fowler one can introduce a thermodynamical potential for which μ (or Z), V, T are the natural independent variables, namely PV; for, from (6·23) and (6·24) it follows that

$$d(PV) = PdV + SdT + Nd\mu. \tag{6·25}$$

For R finally we obtain a formula similar to (6·22). If we introduce the free energy per molecule $a = A/N$, then from (4·3) one gets with (6·4)

$$R = \left(\frac{2\pi mkT}{h^2}\right)^{\frac{3}{2}} e^{a/kT}, \tag{6·26}$$

$$a = A/N.$$

7. Triple points and the critical point

If the singularity $\overline{Y}$ is on the circle of convergence of $G_0(y, \beta)$, we can apply the theorem of Cauchy and Hadamard; with (6·19) we have

$$\frac{1}{\overline{Y}} = v_s = \overline{\lim_{\nu \to \infty}} \left(| \beta_\nu | \right)^{1/\nu},$$

where $\overline{\lim}$ denotes the upper limit of the limit points of the β_ν. If there are several limit points we can relate to each of them a partial series β_{ν_i} so that

$$\left.\begin{aligned} v_s &= 1/\overline{Y} = \max \beta_i \\ \beta_i &= \lim_{\nu_i \to \infty} (|\beta_{\nu_i}|)^{1/\nu_i}. \end{aligned}\right\} \tag{7·1}$$

Since $d\beta_\nu/dT$ is continuous, it follows that in general dv_s/dT is continuous. A discontinuity will occur if a temperature T_3 exists such that the greatest

β_t is the limit of two different partial series above and below T_3. The vapour pressure can be obtained from (5·19):

$$P_s = \frac{kT}{v_s}\left[1 - \Sigma\frac{\nu}{\nu+1}\beta_\nu/v_s^\nu\right], \quad \bar{Y} = \frac{1}{v_s} \tag{7·2}$$

dP_s/dT therefore is in general continuous except at the temperatures T_3 which we may therefore identify with the triple points.

If $\bar{Y}$ is not on the circle of convergence, we obtain similar formulae by applying the theorem of Cauchy and Hadamard to the series representing the analytical continuation of $G_0(y, \beta)$.

If at the triple point there are two crystal structures in equilibrium with the gaseous phase, (7·1) has a simple physical meaning. For the contributions to the integrals β_ν for large values of ν come mostly from parts of phase space in which the molecules have some ordered arrangement. It appears from (7·1) that the partial series β_{ν_t} can directly be related to various crystal structures.

Finally we shall consider the critical temperature T_c where the condensation volume v_s itself is singular. It is obvious that this is impossible if v_s is determined by (7·1). We must assume that there is a temperature region below the critical temperature in which $\bar{Y}$ is determined by (6·20).* In this case v_s may be singular, if we assume that (6·20) has two roots on the positive axis which coincide for $T = T_c$ and vanish into the complex plane for $T > T_c$. In this case we have

$$G_1(\bar{Y}, \beta) = 1, \quad G_2(\bar{Y}, \beta) = 0, \quad T = T_c \tag{7·3}$$

and therefore from (6·21)

$$\frac{dP}{dV} = 0, \quad \frac{d^2P}{dV^2} = 0, \quad T = T_c. \tag{7·4}$$

The considerations about the critical point must however be taken with reserve, as is apparent if we consider the question of how the volume of the condensed phase may be defined in the present theory. One might be tempted to identify the second root of $G_1(y, \beta) = 1$ with this volume. However, there is no justification for doing so. Indeed there is no possibility of defining the volume of the condensed phase in the present theory. We must assume that the condensed phase is in that region where the theory breaks down, i.e. where b_l and β_ν are dependent on the volume.* For this reason the equations for the critical temperature may need modification.

---

* Mayer and Harrison (1938) have identified this region with the region between the critical temperature and the temperature where the meniscus between the liquid and the solid phase disappears.

CONCLUSIONS

We have shown that the exact evaluation of the partition function for a system of equal particles exerting central forces leads to the complete thermodynamics of a gas, including the phenomenon of condensation. One physical assumption had to be made from the beginning, namely that the coefficients of the partition function, which Mayer has introduced, the irreducible integrals β_ν, do not depend on the volume. This is certainly true for large volumes, but not for the condensed phase. In spite of this fact the theory explains not only condensation but indicates even the possibility of the existence of different condensed phases. It appears, at first sight, astonishing that it is possible to obtain such far reaching results without knowing anything about the irreducible integrals β_ν. We have, however, made full use of the fact that the partition function is positive, and this obviously imposes rather severe conditions on the β_ν. As Mayer has discovered, the phenomenon of condensation is connected with the singularities of the series, which represent the thermodynamical properties, and as these series have coefficients depending on the β_ν, the positivity condition mentioned is just sufficient to determine the singularity and thereby the condensation point.

REFERENCES

Born, M. 1937 *Physica*, **4**, 1034.
Mayer, J. E. 1937 *J. Chem. Phys.* **5**, 67.
Mayer, J. E. and Ackermann, Ph. G. 1937 *J. Chem. Phys.* **5**, 74.
Mayer, J. E. and Harrison, S. F. 1938 *J. Chem. Phys.* **6**, 87, 101.
Uhlenbeck, G. E. and Kahn, B. 1938 *Physica* (in the press).
Ursell, H. D. 1927 *Proc. Cambr. Phil. Soc.* **23**, 685.

NOTATION

Q	partition function.
Q_τ	factor of Q due to the potential energy multiplied with $N!$
$A = E - TS$	Helmholtz' free energy.

---

* Mayer and Harrison (1938) have developed the expressions which have to replace the functions $G_\lambda(z, vb)$ in the case that the b_l depend on the volume. It should be investigated whether these formulae lead to a description of the condensed phase.

$F = A + PV$	Gibb's thermodynamic potential.
$v = V/N$	volume per molecule.
b_l	cluster integrals.
β_ν	irreducible integrals.

$$F(M, I, x_1, x_2 \ldots) = F(M, I, x) = \sum_{\substack{m_l \\ \Sigma l m_l = M}} \prod_{l} \frac{(Ix_l)^{m_l}}{m_l!}.$$

$$G_\lambda(\xi, x_1, x_2 \ldots) = G_\lambda(\xi, x) = \sum_{\nu=1}^{\infty} \nu^\lambda x_\nu \xi^\nu.$$

$$H_\lambda^m(r, x_1, x_2 \ldots) = H_\lambda^m(r, x) = \sum_{N=1}^{\infty} \frac{F(N-m, N, x)}{N^\lambda} r^N.$$

The symbols r, z, y are reserved for the following series:

(*a*) $H_\lambda^0(r, vb)$.

(*b*) $G_\lambda(z, vb)$.

(*c*) $G_\lambda(y, \beta)$.

$R =$	radius of convergence of (*a*). (R is also used for the gas constant but there is no possibility of mistaking these two quantities.)
Z, Y	are quantities corresponding to R.
$\bar{Z} =$	nearest singularity of the series (*b*) on the positive axis.
$\bar{Y}$	is a quantity corresponding to $\bar{Z}$.

An accurate determination of the range-distribution curve of the radioactive α-particles from ^{8}Li

By C. L. Smith, B.A., *Sidney Sussex College, Cambridge,*
and W. Y. Chang, B.A., *Jesus College, Cambridge*

(*Communicated by J. D. Cockcroft, F.R.S.—Received* 12 *March* 1938)

The bombardment of lithium by deuterons has been shown by Crane, Delsasso, Fowler and Lauritsen (1935) and by Rumbaugh and Hafstad (1936) to lead to the formation of a radioactive product which emits β-rays and has a half life of 0·88 sec. Recently, Lewis, Burcham and Chang (1937) showed that this emission of β-rays is accompanied by α-particle emission which decays with the same period.

These two disintegrations may be explained by the following chain of reactions in which the radioactive ^{8}Li nucleus with a half life of 0·88 sec. disintegrates into two α-particles and an electron and neutrino:

$$^7\mathrm{Li} + {}^2\mathrm{D} = {}^8\mathrm{Li} + {}^1\mathrm{H},$$

$$^8\mathrm{Li} = {}^4\mathrm{He} + {}^4\mathrm{He} + \bar{e} + \nu.$$

It is customary, however, to postulate the intermediate formation of an excited ^{8}Be nucleus of very short life which disintegrates into two α-particles:

$$^8\mathrm{Li} = {}^8\mathrm{Be} + \bar{e} + \nu,$$

$$^8\mathrm{Be} = {}^4\mathrm{He} + {}^4\mathrm{He}.$$

To explain their results on the disintegration of boron into three α-particles under proton bombardment Dee and Gilbert had suggested (1936) that an excited ^{8}Be was formed as a short-lived intermediate product:

$$^{11}\mathrm{B} + {}^1\mathrm{H} = {}^8\mathrm{Be}^* + {}^4\mathrm{He},$$

$$^8\mathrm{Be}^* = {}^4\mathrm{He} + {}^4\mathrm{He}.$$

They postulated an energy of excitation of the excited ^{8}Be of about 3×10^6 e-volts with a "half width" of 1×10^6 e-volts. If the ^{8}Be produced according to the lithium reaction had this same excitation energy one would expect its disintegration to give rise to a heterogeneous group of α-particles, most of which would have ranges between approximately 1·0 and 2·5 cm. Actually Lewis, Burcham and Chang found that the majority had ranges less than 1·5 cm., but a few had ranges up to 5 cm., and though their experiments were rather inaccurate, their results suggested that the number decreased continuously from 1·0 to 5·0 cm. range. They did not investigate the curve for α-particle ranges of less than 0·8 cm. ($1\cdot5\times10^6$ e-volts). If the ^{8}Be formed in this reaction has an excited state at about 3×10^6 e-volts the yield of α-particles should fall off below 0·8 cm. ($1\cdot5\times10^6$ e-volts). Since that time Rumbaugh, Roberts and Hafstad (1937) and Fowler and Lauritsen (1937) have published number-range distribution curves for these α-particles. Rumbaugh, Roberts and Hafstad used counters and Fowler and Lauritsen used a cloud chamber, and the results of these two experiments are not in agreement. Both groups give curves which extend (in particle range) from about 0·5 up to 5 cm. The aim of the present paper is to describe experiments using counters which, we think, give a more accurate distribution curve for these α-particles. The portion of the distribution curve which we have investigated extends from 0·68 cm. up to approximately 6·9 cm.

METHOD

In the first part of the experiments, for the determination of the curve from 0·68 up to 5 cm., thick Li targets were bombarded with about 100 μA of deuterons at a bombarding voltage of 550 kV. The investigation of the distribution of α-particles with ranges greater than 5 cm. is more difficult owing to the small yield at these ranges. For this portion of the curve therefore the recently erected high-voltage apparatus in the Cavendish Laboratory was used. With this apparatus a deuteron beam of about 50 μA at 1×10^6 e-volts was available. This increased greatly the yield of α-particles.

The lifetime being 0·88 sec., intermittent activation and counting for periods of the order of a few seconds each is required. To enable this to be done "continuously", a rotating shutter S was placed in the path of the beam (see fig. 1). The shutter consisted of an aluminium disk 7·8 cm. in diameter and 0·3 cm. thick. A sector of 120° was cut out of the disk. This disk was rotated by means of a small induction motor M (as used in "Ferranti" clocks). The spindle of the motor rotated six times per second, and the disk was driven through a set of reduction gears, G, so that it rotated once in 6 sec. The motor and the reduction gears were placed in a cylindrical brass box above the tube containing the target and the driving spindle projected into the target box T. The shutter S was placed directly above one of the targets L of Li_2O which was deposited on a piece of copper foil and mounted on a target head *H* which could be water-cooled. Four targets could be rotated into the path of the beam by means of the ground joint *J*.

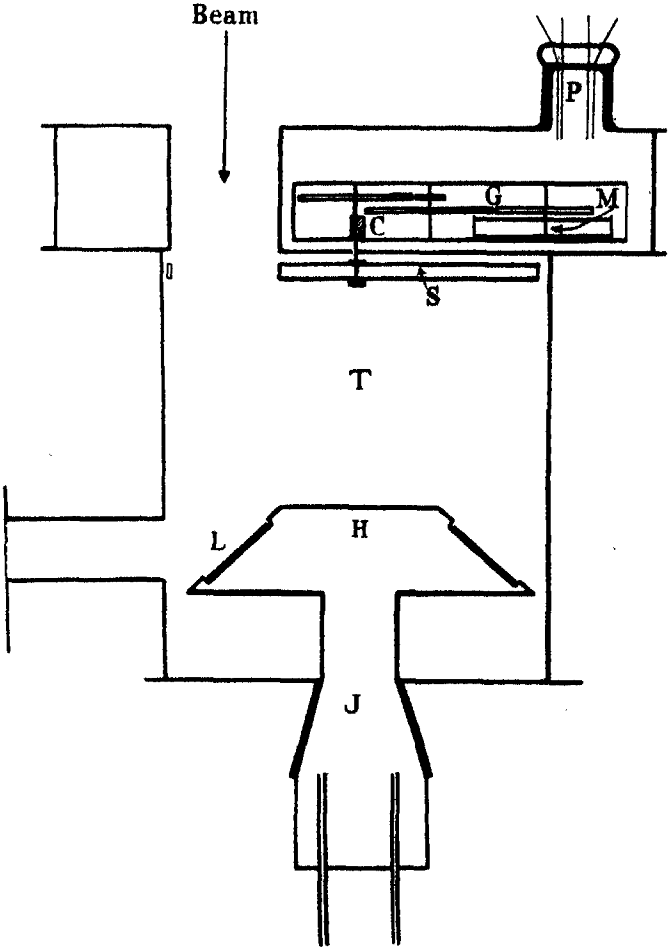


FIG. 1. Apparatus.

During each revolution of the shutter, lasting 6 sec., the beam bombarded the target for 2 sec. and was cut off for 4. The particles from the radioactive body formed during the bombardment were counted during the periods of 4 sec. when the beam was cut off by the shutter. In order not to count the particles resulting from the following reactions:

$$^{6}\mathrm{Li} + ^{2}\mathrm{D} = ^{4}\mathrm{He} + ^{4}\mathrm{He}$$

$$^{7}\mathrm{Li} + ^{2}\mathrm{D} = ^{4}\mathrm{He} + ^{4}\mathrm{He} + ^{1}n$$

the amplifiers were short-circuited during the period of bombardment and the counting was carried out only during the time when the beam was not falling on the target. The switch used for shorting the amplifiers was operated by a cam *C* on the spindle driving the shutter. The leads to the motor and the shorting switch were taken out through a glass pinch *P*. The whole of the apparatus was sealed on to the bottom of the accelerating tube with Apiezon Q sealing compound and evacuated during the experiment.

Two flat mica windows were arranged opposite the target to allow the particles to pass out into the air. The planes of these windows were at right angles to each other, and opposite each window an ionization chamber was placed to count the particles. Each ionization chamber was connected to a linear amplifier and a "scale-of-two" counter. Opposite one window an ionization chamber was placed which was kept at a fixed distance from the target during the experiment. Opposite the other window a differential chamber was placed and the range-distribution curve was determined by moving this chamber towards and away from the target. With this arrangement it was not necessary to know the target current, since the ratio of the count in the differential chamber to that in the fixed chamber gave the ratio of the number of particles with the range determined by the position of the differential chamber to that at a fixed range given by the position of the fixed chamber. Thus to determine the range-distribution curve it was only necessary to plot this ratio against the range as determined by the position of the differential chamber, troubles due to fluctuations of the bombarding current being thus entirely eliminated. The mica window in front of the differential chamber had an air equivalent of 0·27 cm. The particles all lay within a cone making an angle of 10° with the direction at right angles to the bombarding beam. The window opposite the fixed chamber was of large aperture, since exact collimation was not necessary in this case, and was of 1·15 cm. air equivalent. The above arrangements were used for the investigation of the distribution of ranges up to 5 cm.

The yield of α-particles at higher ranges (*ca.* 6·7 cm.) is only 10^{-4} of that at a range of 0·7 cm.; thus when the range distribution beyond 5 cm. was examined it was necessary to sacrifice the good collimation to some extent in order that a reasonable number of α-particles should be counted. A mica window subtending a larger angle at the target was therefore used and the bombarding voltage was increased from 550 to 900 kV.

The differential chamber counted all particles within a range of 0·27 cm. Thus each point in the distribution curve corresponds to a count of the number of particles within 0·27 cm. of the range indicated as abscissae. At each setting of the chamber, approximately 1000 particles were counted in the differential chamber for ranges from 0·68 up to 5·5 cm. Beyond this range usually 300 particles were counted at each point except in the last 0·3 cm. of the curve, where only 20–60 particles were counted per point.

RESULTS

Fig. 2 shows the number-range curve which we have obtained for these α-particles. The ordinates are plotted on a log scale and the abscissae are

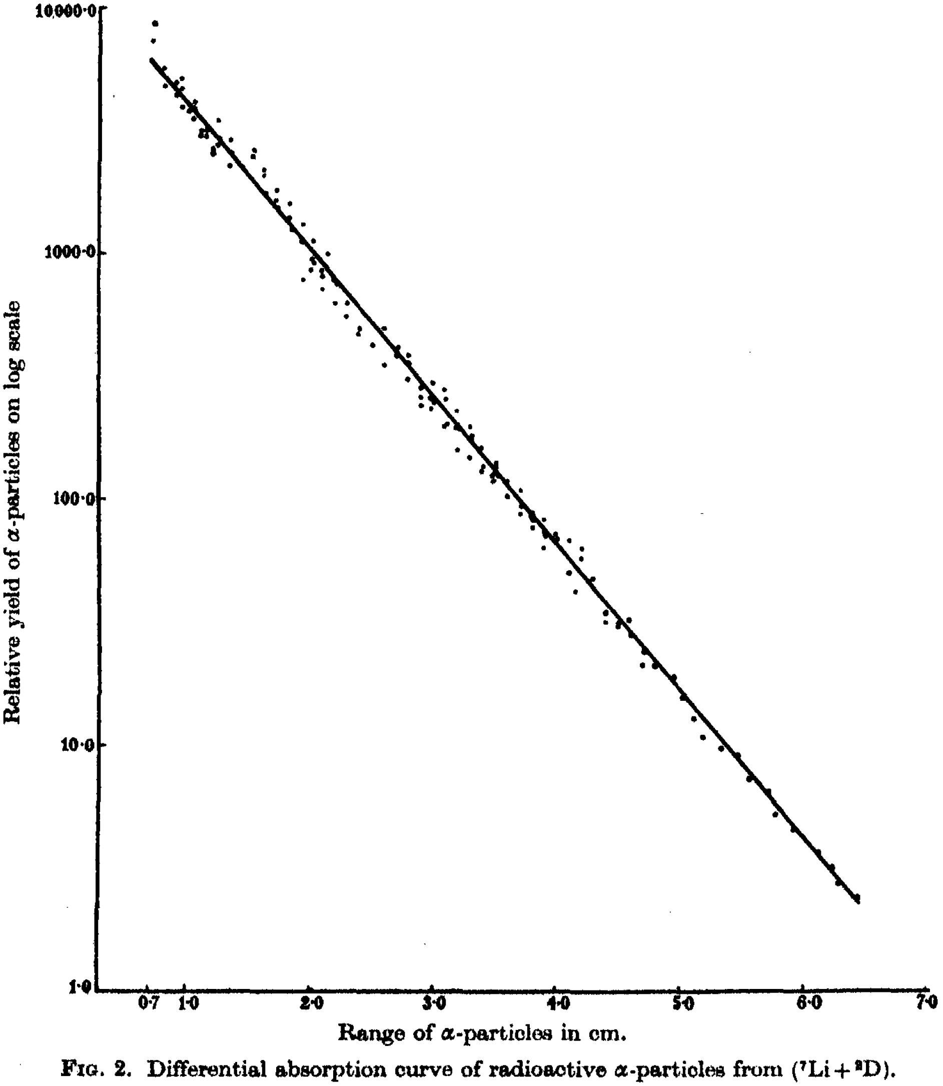


FIG. 2. Differential absorption curve of radioactive α-particles from (^{7}Li + ^{2}D).

plotted directly in ranges. This curve extends from 0·67 up to 6·5 cm. range. It appears from this curve that there is a linear relation between the logarithm of the number of α-particles and their range.

Fig. 3 shows three curves. Curve 1 shows the distribution near the end-point of the α-particles under discussion, while curve 2 shows the distribution near the end-point of the α-particles from the reaction

$$^{7}\mathrm{Li} + {}^{2}\mathrm{D} = {}^{4}\mathrm{He} + {}^{4}\mathrm{He} + {}^{1}n. \qquad (a)$$

Curve 3 shows the distribution near the end-point of the α-particles from the reaction

$$^{7}\mathrm{Li} + {}^{1}\mathrm{H} = {}^{4}\mathrm{He} + {}^{4}\mathrm{He}. \qquad (b)$$

Curves 2 and 3 were obtained at a bombarding voltage of 300 kV.

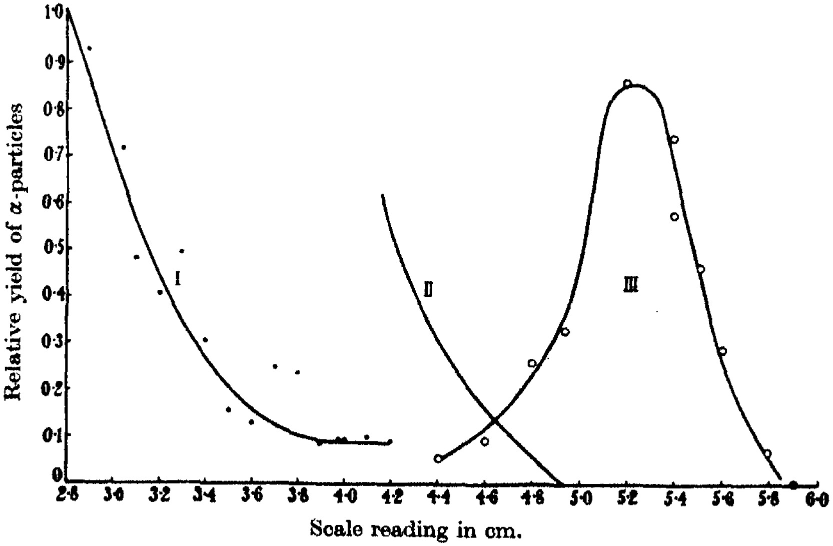


Fig. 3. Distribution curves.

I, distribution near the end-point of radioactive α-particles from (Li + D).

II, distribution near the end-point of continuous distribution of α-particles from Li + D $\rightarrow 2\alpha + n$.

III, distribution near the end-point of 8·4 cm. group of α-particles from (Li + ^{1}H).

From curve 1 we estimate that the end-point of the continuous distribution of the radioactive α-particles is at a scale reading of 3·9 ± 0·1 cm., and to determine its range we have compared it with curves 2 and 3. Oliphant, Kempton and Rutherford (1935) have shown that, at a deuteron energy of $0{\cdot}16 \times 10^{6}$ e-volts, the continuous distribution of particles from reaction (a) have a maximum range of 7·8 cm. and that, for a bombarding proton energy of $0{\cdot}19 \times 10^6$ e-volts, the group of α-particles from reaction (*b*) have a mean range of $8{\cdot}29 \pm 0{\cdot}03$ cm. These two curves give for the maximum range of the radioactive particles $6{\cdot}8 \pm 0{\cdot}1$ and $6{\cdot}95 \pm 0{\cdot}1$ cm. respectively. Since the end-point of a group of particles can be determined with greater accuracy than that of a continuous distribution, we think that the range given by comparison with the 8·29 cm. group is more accurate and suggest that the α-particles from radioactive ^{8}Li have a maximum range of $6{\cdot}95 \pm 0{\cdot}1$ cm., i.e. an energy of $7{\cdot}75 \pm 0{\cdot}05 \times 10^6$ e-volts.

DISCUSSION

The curve we have obtained shows that radioactive ^{8}Li emits α-particles with a continuous range of energies from $1{\cdot}2 \times 10^6$ e-volts (0·65 cm.) up to $7{\cdot}75 \times 10^6$ e-volts (6·95 cm.), and that in the range from 0·65 up to 6·5 cm. there is a linear relation between the logarithm of the number of α-particles and their range. If Ndx are the number of α-particles with ranges between x and $x + dx$ this means that

$$N = N_0 e^{-\alpha x}.$$

The characteristic of such a curve is the range interval in which the value of N diminishes by a factor $1/e$ and from the curve we see that this is 0·745 cm.

A comparison of our distribution curve with those of other workers (fig. 4) shows several interesting differences. Whereas previously α-particles of range greater than about 5 cm. had never been reported we have extended the observations to a range of 6·95 cm. and have shown that this is the end-point of the distribution of radioactive α-particles. This enables us to give a lower limit for the mass of ^{8}Li. Plotting our curve and those of other workers on the same graph we find that the curve given by Rumbaugh, Roberts and Hafstad (1937) agrees with ours down to the lowest observed point of the range 0·65 cm. and up to 5 cm. the limit of their observations. On the other hand, the curve obtained by Fowler and Lauritsen (1937) shows a broad maximum at an energy of $1{\cdot}3 \times 10^6$ e-volts. We can obtain no evidence for this maximum. However, owing to the decrease in probability of the mutual separation of two α-particles, i.e. the disintegration of a ^{8}Be nucleus, for low energies of the emitted α-particles one would expect the yield of α-particles to fall off for very low energies but according to our measurements this does not occur at an energy greater than $1{\cdot}2 \times 10^6$ e-volts (0·65 cm. range).

Consider the energy balance in the equation representing the radioactive disintegration

$$^8\mathrm{Li} = {}^4\mathrm{He} + {}^4\mathrm{He} + \bar{e} + \nu + Q.$$

On the assumption that the emission of α-particles of maximum energy corresponds to the emission of the electron and neutrino with zero energy we can calculate the mass of ^{8}Li, since this will be the mass of two α-particles plus twice the maximum observed energy of the α-particles. This gives

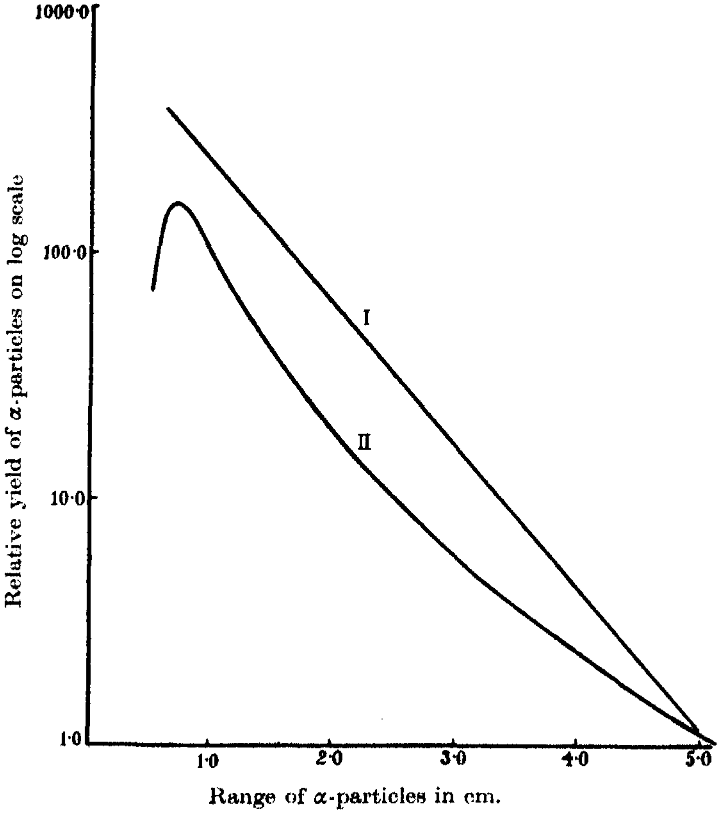


FIG. 4. Differential absorption curves of the radioactive α-particles from (^{7}Li + ^{2}D).

I, Smith and Chang, and Rumbaugh, Roberts and Hafstad.
II, Crane and Lauritsen.

8·0246 for the mass of ^{8}Li. However, our assumption may be wrong and α-particles of maximum energy may correspond to the emission of the $(\bar{e}+\nu)$ with a certain amount of energy. In this case one has to consider the above mass of ^{8}Li as a lower limit. From consideration of the energy balance in the equation

$$^7\mathrm{Li} + {}^2\mathrm{D} = {}^8\mathrm{Li} + {}^1\mathrm{H} + Q_1,$$

one can obtain an estimate of the upper limit of the mass ^{8}Li. We have observed that ^{8}Li can be formed at as low a bombarding voltage as 370 kV. The proton must carry some energy away, but assuming that the proton receives no energy at all we can estimate the mass of ^{8}Li, and this estimate will be an upper limit, since if we take into account proton energy it can only decrease the estimate of the mass and not increase it. By this means we obtain the upper limit of the mass of ^{8}Li to be 8·0252, and we can now definitely state that the mass of ^{8}Li lies between the limits 8·0246 and 8·0252.

That the proton is certainly emitted with only a small energy has been shown by Rumbaugh, Roberts and Hafstad who bombarded Li with deuterons at a bombarding voltage of 850 kV. and were unable to find any protons with a range greater than 2·7 cm. which could be ascribed to the above reaction. This suggests that Q_1, the energy release in the above reaction, is certainly less than $0{\cdot}27 \times 10^6$ e-volts. From our value for the lower limit to the mass of ^{8}Li we find by substitution in equation (2) that the proton energy cannot exceed $0{\cdot}25 \times 10^6$ e-volts.

A few interesting conclusions concerning the β-ray spectrum may be drawn from the α-particle distribution curve. Knowing all the masses in the reaction

$$^7\mathrm{Li} + {}^2\mathrm{D} = {}^4\mathrm{He} + {}^4\mathrm{He} + \bar{e} + \nu + {}^1\mathrm{H} + Q_2,$$

we can show that the energy release is $15{\cdot}6 \times 10^6$ e-volts, and since the proton is released with less than $0{\cdot}25 \times 10^6$ e-volts energy the energy release in the reaction

$$^8\mathrm{Li} = {}^4\mathrm{He} + {}^4\mathrm{He} + \bar{e} + \nu + Q$$

cannot differ by more than $0{\cdot}27 \times 10^6$ e-volts from $15{\cdot}6 \times 10^6$ e-volts. Now considering the division of the available energy between the α-particles and the $(e + \nu)$ it is seen that if in the disintegration of a single ^{8}Li nucleus, the $(e + \nu)$ take away energy E, then the energy to be divided between the two α-particles is $(Q - E) = (15{\cdot}6 - E) \times 10^6$ e-volts. Conversely, α-particles of this energy should give rise to a simple β-ray spectrum of maximum energy E, and therefore, as has been pointed out by others, a continuous distribution of α-particles would indicate that the β-ray spectrum was a complex one consisting of the superposition of an infinite number of simple β-ray spectra. On this picture one might expect the α-particles to extend from zero energy in which the $(\bar{e} + \nu)$ took all the energy up to a maximum of $15{\cdot}6/2 \times 10^6$ e-volts in which the $(\bar{e} + \nu)$ are emitted with zero energy. The first point we notice is that since our α-particle curve does extend up to an energy of $15{\cdot}6/2 \times 10^6$ e-volts $= 7{\cdot}8 \times 10^6$ e-volts, then apparently an $(\bar{e} + \nu)$ can be emitted in nuclear processes with zero energy.

On this picture one would also expect the β-ray spectrum to extend up to a maximum of $15 \cdot 6 \times 10^6$ e-volts. Experimental determination of the curve indicates that the upper limit is at about 12×10^6 e-volts. This suggests that the α-particles emitted in this process have always an energy greater than $\frac{15 \cdot 6 - 12}{2} \times 10^6$ e-volts $= 1 \cdot 8 \times 10^6$ e-volts. Since our curve for the α-particle distribution extends well below this value this conclusion is unacceptable, and we suggest that the β-rays of maximum energy have not yet been observed.

In conclusion we would like to thank Dr Lewis for many valuable suggestions and help during the first half of the experiment. We are also indebted to Mr P. I. Dee and Dr W. E. Burcham for similar assistance during the latter part of the work, which was carried out in the New High Tension Laboratory, of the Cavendish Laboratory. One of us (C. L. S.) would like to thank both the Department of Scientific and Industrial Research and the Governing Body of Sidney Sussex College for maintenance grants.

SUMMARY

The paper describes a method for determining the number-range distribution curve for the α-particles from radioactive lithium. These α-particles are shown to consist of a continuous distribution extending up to a maximum of $6 \cdot 95 \pm 0 \cdot 1$ cm. ($7 \cdot 75 \pm 0 \cdot 05 \times 10^6$ e-volts). The shortest range observed was $0 \cdot 65$ cm. ($1 \cdot 2 \times 10^6$ e-volts), and in the region from $0 \cdot 65$ up to $6 \cdot 6$ cm. there is a linear relation between the logarithm of the number of α-particles and their range. Comparison is made with the distribution curves obtained by other workers and it is also shown that the upper limit of the energy of the α-particles explains the failure to observe the protons emitted in the reaction

$$^7\mathrm{Li} + {}^2\mathrm{D} = {}^8\mathrm{Li} + {}^1\mathrm{H}.$$

Upper and lower limits for the mass of ^{8}Li are calculated from the data and shown to respectively 8·0253 and 8·0246.

REFERENCES

Crane, Delsasso, Fowler and Lauritsen 1935 *Phys. Rev.* **47**, 971.
Dee and Gilbert 1936 *Proc. Roy. Soc.* A, **154**, 279.
Fowler and Lauritsen 1937 *Phys. Rev.* **51**, 1103.
Lewis, Burcham and Chang 1937 *Nature, Lond.*, **139**, 24.
Oliphant, Kempton and Rutherford 1935 *Proc. Roy. Soc.* A, **149**, 406.
Rumbaugh and Hafstad 1936 *Phys. Rev.* **50**, 681.
Rumbaugh, Roberts and Hafstad 1937 *Phys. Rev.* **51**, 1106.

On the theory of scattering of light

By Hans Mueller*

Cavendish Laboratory, Cambridge

(*Communicated by R. H. Fowler, F.R.S.—Received* 16 *March* 1938)

THE KRISHNAN EFFECT

In a series of recent investigations R. S. Krishnan (1934–8) demonstrated the existence of a new effect which will be called the Krishnan effect. It relates to the state of polarization of the light scattered by certain liquid or solid media in directions normal to the incident beam. To describe the effect let us denote with Π the plane parallel to the direction of observation and to that of the incident beam. Since in the experiment this plane is usually horizontal we denote by H the intensity of those scattered light components which vibrate parallel to this plane, and by V those vibrating normal to Π. In a similar manner subscripts h or v indicate whether the incident light vibrates parallel or normal to the plane. We distinguish therefore (see fig. 1) the four light components H_h, H_v, V_h and V_v. Following Krishnan the depolarizations are defined by

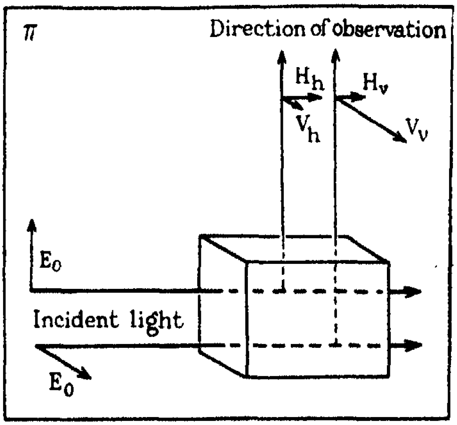


Fig. 1. Schematic representation of the four components of the scattered light.

$$\rho_h = V_h/H_h, \quad \rho_v = H_v/V_v, \quad \rho_u = (H_h+H_v)/(V_h+V_v).$$

---

* Fellow of the John Simon Guggenheim Memorial Foundation, on leave of absence from Massachusetts Institute of Technology, Cambridge, Mass.

ρ_u is the depolarization for natural incident light. For most liquids the observations give, in agreement with the theories of temperature scattering, $H_h = V_h = H_v$, hence $\rho_h = 1$, $\rho_u = 2\rho_v/(1+\rho_v)$. The Krishnan effect is the observation that in a number of liquid and solid systems

$$\rho_h = V_h/H_h \neq 1, \tag{1}$$

and

$$V_h = H_v. \tag{2}$$

Krishnan has called (2) the reciprocity relation. All observations have given $\rho_h < 1$, but none of the present theories exclude the possibility that ρ_h may assume values larger than 1.

The Krishnan effect has been observed for colloids (Krishnan 1937), proteins (Subbaramaija 1934, 1935; Ramaiah 1937) and emulsions, for rubber solutions in various solvents (Gehmann and Field 1937), for oils, for mixtures of liquids if the mixing ratio and the temperature are near the critical point (Krishnan 1934*d*, 1935), for the associated liquids of formic and acetic acid, and for glasses (Krishnan 1936*b*). For colloidal solutions the effect can be explained with Mie's (1908) theory, if it is assumed that the particles are not spherical and have linear dimensions of the order of magnitude of the wave-length of light. Krishnan therefore concluded that all other substances which show the effect have a "colloid-like" structure, i.e. in liquids it is due to the formation of molecular clusters. Cluster formation will indeed introduce additional density variations in the liquid which have not been taken into consideration in the theories of temperature scattering. These variations give rise to a new type of scattering which Krishnan calls "structure scattering". Ornstein and Van Cittert (1935) proposed that this type of scattering exists even in ordinary liquids and is responsible for the unmodified frequency component of the scattered light. The older work of Ornstein and Zernicke (1926) and of Placzek (1930) on light scattering near the critical point cannot account for the Krishnan effect, because, as Gans has pointed out, it deals only with density scattering and neglects orientation scattering. Gans (1936) therefore developed a new theory of scattering by molecular clusters. He finds that the Krishnan effect requires the existence of optically anisotropic clusters which are strongly aspherical in shape.

Gans' theory is not an entirely satisfactory solution of the problem, because it involves the assumption that the clusters are small in comparison with the wave-length of light. In a second paper Gans (1937) questioned the validity of Krishnan's reciprocity relation. However, the following consideration shows that his objections are not valid. In Gans' notation V_h is the mean square of the quantity (equation (10) of Gans' (1936) paper)

$$f_z' = (g_1\alpha_3\alpha_2 + g_2\beta_3\beta_2 + g_3\gamma_3\gamma_2) \\ \times\left[1 - \frac{1}{2}\frac{k^2}{V}\{(\alpha_1+\alpha_2)^2\theta_1 + (\beta_1+\beta_2)^2\theta_2 + (\gamma_1+\gamma_2)^2\theta_3\}\right],$$

where g_i, θ_i, k^2 and V are constants and the average must be formed by letting the direction cosines α_i, β_i, γ_i assume all values compatible with arbitrary rotations of the clusters. If the calculation of Gans is carried through for an incident light beam which vibrates in the v direction ($E = E_z$ in Gans' notation), it can easily be seen that H_v is the mean square of

$$f_x' = (g_1\alpha_1\alpha_3 + g_2\beta_1\beta_3 + g_3\gamma_1\gamma_3) \\ \times\left[1 - \frac{1}{2}\frac{k^2}{V}\{(\alpha_1+\alpha_2)^2\theta_1 + (\beta_1+\beta_2)^2\theta_2 + (\gamma_1+\gamma_2)^2\theta_3\}\right].$$

We notice that f_x' can be obtained from f_z by exchanging the subscripts 1 and 2. Such an exchange can have no influence on the mean square values of the two quantities. Hence, contrary to Gans' second paper, it follows from his own theory that Krishnan's reciprocity relation $V_h = H_v$ must be valid.*

Gans' theory can give no explanation for the fact that ρ_h is always found to be smaller than unity. It rather would lead one to expect that for most substances ρ_h should be larger than 1, because $\rho_h < 1$ requires somewhat artificial assumptions concerning the geometrical form and the direction of maximum fluctuation of the optical anisotropy of the clusters (Vrkljan and Katalinić 1936).

More serious objections against the cluster hypothesis arise, if it is applied to explain the Krishnan effect in glasses. The recent X-ray investigations of Warren (1937) have eliminated the possibility for the existence of larger crystalline or quasi-crystalline units in glasses. Warren's non-repeating network structure of glasses leaves no place for Krishnan's "molecular aggregates, the size of which is not exceedingly small compared to the wave-length of light".

---

* In a letter Professor Gans has informed the writer that he arrived at the same conclusion, by the same method. I am indebted to Professor Gans for the permission to publish the above consideration.

KRISHNAN EFFECT IN TEMPERATURE SCATTERING

It has generally been assumed that the light scattering which is produced by the temperature fluctuations could not give rise to a Krishnan effect in an isotropic and homogeneous liquid or solid. We should like to point out that this assumption is not necessarily correct. We can show that the temperature scattering in homogeneous solids must always show the Krishnan effect and that it might exist even for ordinary liquids. This statement does not, however, imply that Krishnan's observation can be explained without assuming structure scattering.

Since its inception the continuum theory of light scattering in liquids, as formulated in the fundamental paper of Einstein, has required only two modifications. The first is due to Gans, who, in order to explain the depolarization of the scattered light, introduced fluctuations of the optical anisotropy. The second change, due to Brillouin, introduced Debye's treatment of the temperature fluctuations and led to the discovery of the frequency change due to Doppler effect. Until recently these two modifications have not been combined. Brillouin's theory considers only density fluctuations, and the anisotropy fluctuations in liquids have not been treated by Brillouin's method.

Both modifications have one common characteristic. They borrow ideas from the field of crystal physics and apply them successfully to explain properties of liquids. It is therefore natural that, to understand the Krishnan effect, we should turn to the problem of light scattering in solids.

In their theory of the temperature scattering in crystals Leontowitsch and Mandelstamm (1932) have combined the ideas of Gans and of Brillouin. This interesting paper leads to a series of conclusions which the authors apparently did not realize. They give the following equation for the transverse scattering in an isotropic solid:*

$$I = \frac{\pi^2 n^8}{\lambda_0^4} \frac{kT}{2} \left[\frac{p_{12}^2}{c_{11}} + \frac{p_{44}^2}{c_{11}} + \frac{p_{44}^2}{c_{44}} \right]. \tag{3}$$

I is the intensity of the light scattered by 1 cm.^{3} for natural incident light of intensity 1, λ_0 is the wave-length of light, n is the refractive index, c_{ik} are the elastic constants and p_{ik} are Pockel's elasto-optical constants. For isotropic solids $2c_{44} = c_{11} - c_{12}$ and $2p_{44} = p_{11} - p_{12}$. If a solid shows no photoelastic effects, i.e. if it cannot be made optically anisotropic, $p_{44} = 0$.

---

* This expression is obtained from equation (16) of Leontowitsch and Mandelstamm (1932) by choosing $r = 1$ and by taking into account that for isotropic solids $c_{11} - c_{12} = 2c_{44}$.

Hence the first term p_{12}^2/c_{11} is primarily due to density scattering, and the other terms are due only to anisotropy or orientation scattering.

In an earlier paper Gans (1925) derived another expression for the scattering in isotropic solids which differs from (3) in so far as the expression in brackets takes the value

$$\left[\frac{p_{12}^2}{c_{11}}+\frac{4}{3}\frac{p_{11}p_{44}}{c_{11}}+\frac{8}{5}\frac{p_{44}^2}{c_{11}}+\frac{26}{15}\frac{p_{44}^2}{c_{44}}\right]. \tag{4}$$

The difference arises from the fact that in the theory of Gans the fluctuations in neighbouring volume elements are assumed to be independent of each other, while Leontowitsch and Mandelstamm represent the temperature motion by a set of standing elastic waves. Comparison of equations (3) and (4) shows that this fundamental difference does not alter the part which is due to density scattering, except for a small factor which is due to the circumstance that in equation (3) c_{ik} are the adiabatic elastic constants while in equation (4) they are the isothermal constants. This result is not unexpected because Brillouin (1922) arrived at the same conclusion. More important is the result that the two theories give different results for the scattering which is due to the fluctuation of the optical anisotropy. It signifies that *the laws for the depolarization of the scattered light must be altered if one takes into account that not only the density fluctuations, but also the fluctuations of the optical anisotropy are not independent in neighbouring volume elements.* I shall show that this change gives rise to the Krishnan effect.

Leontowitsch and Mandelstamm obtain their result (equation (3)) through the addition of four terms, $\overline{\Gamma_{12}^2}+\overline{\Gamma_{13}^2}+\overline{\Gamma_{23}^2}+\overline{\Gamma_{33}^2}$. A careful study of their derivation reveals that these four quantities correspond to Krishnan's components, as follows:*

$$\left.\begin{aligned} V_v &= \overline{\Gamma_{33}^2} = p_{12}^2/c_{11}, \quad & V_h &= \overline{\Gamma_{23}^2} = \tfrac{1}{2}p_{44}^2/c_{44}, \\ H_h &= \overline{\Gamma_{12}^2} = p_{44}^2/c_{11}, \quad & H_v &= \overline{\Gamma_{13}^2} = \tfrac{1}{2}p_{44}^2/c_{44}, \end{aligned}\right\} \tag{5}$$

where a common factor of proportionality has been omitted.

Hence it follows that temperature scattering in solids must satisfy the reciprocity relation $V_h = H_v$ and must show a Krishnan effect

$$\rho_h = V_h/H_h = \tfrac{1}{2}c_{11}/c_{44} = (1-\sigma)/(1-2\sigma), \tag{6}$$

---

* The identification of the Γ_{ik} with Krishnan's components is made with the help of equation (14) of the paper by Leontowitsch and Mandelstamm (1932) and the mean squares are found from their equations (9). Another derivation of the above relations is presented here in a later section (see equation (14)).

where σ is Poisson's ratio. In contrast to Krishnan's results temperature scattering always must give $\rho_h > 1$.

An ideal liquid may be considered as a special case of a solid without form elasticity, $\sigma = 0$, and without strain birefringence, $p_{44} = 0$. For this limiting case equation (3) reduces indeed to the well-known expression for density scattering in liquids. An actual liquid, however, has a finite viscosity and shows *streaming birefringence, though the latter effect is frequently too* small to be observed. In a viscous medium shearing strains and transversal waves can exist, as Brillouin (1936) and more recently Lucas (1937) have emphasized. These authors have pointed out that it is difficult to account for the specific heat of liquids without assuming the existence of transversal waves, and Lucas has given new experimental evidence in support of this hypothesis. The transversal waves "of inertia and viscosity" of Lucas will give a contribution to the scattering of light because they create shearing strains, hence streaming birefringence and fluctuations of the optical anisotropy.

For a viscous liquid the extrapolation to $\sigma = 0$ is therefore not justified and p_{44} does not vanish, but the latter must be replaced by a quantity proportional to Maxwell's constant of streaming birefringence.

Our knowledge of the temperature motion in liquids is too rudimentary to justify an attempt[*] to generalize the theory of Leontowitsch and Mandelstamm to liquids. Still, the above considerations make it appear probable that the Krishnan effect might occur in ordinary liquids. The essential conditions for its existence are a high viscosity and strong streaming birefringence. Furthermore, the fluctuations in neighbouring volume elements cannot be independent. According to equation (6) we should expect that for temperature scattering the depolarization ρ_h should always be larger than unity.

Krishnan repeatedly has pointed out that all substances which show his effect also have an abnormally high viscosity and manifest flow birefringence. Theory and experiment agree in so far as both point to a very close correlation between these three properties. Similarly, the idea of cluster formation *involves the assumption that the fluctuations are not independent in small* volume elements. However, the experimental fact $\rho_h < 1$ is not compatible with the assumption that the effect is due to temperature fluctuations. It is

[*] Since $c_{11}/c_{44} = (V_L/V_T)^2$, where V_L and V_T are the sound velocities for longitudinal and transversal waves, equation (6) can also be written $\rho_h = \frac{1}{2}(V_L/V_T)^2$. Lucas (1937) *has shown that the velocities of his transversal waves depend on their* frequency. If one introduces for V_T the velocity of Lucas' transversal wave which produces scattering by Bragg reflexion one finds a value of ρ_h of the order of magnitude 10^3. This is, of course, quite impossible.

therefore necessary to introduce an additional hypothesis, which will be discussed in the following sections.

SCATTERING OF LIGHT IN GLASSES

It is impossible to account for the observed scattering in glasses by the theory of temperature scattering, for the following three reasons:

(*a*) The observed intensities of the scattered light are about fifty times too large. According to the theories of Gans or of Leontowitsch and Mandelstamm the scattering in solids should be about as strong as in gases at normal pressure. Actually the scattering power of glasses is found to be similar to that of most liquids.

(*b*) Since Poisson's ratio for most glasses (Schaefer, Bergmann and Goehlich 1937) is between $\frac{1}{4}$ and $\frac{1}{5}$ the depolarization ρ_h should, according to equation (6), have values between $\frac{3}{2}$ and $\frac{4}{3}$. Krishnan's measurements on seventeen glasses consistently give values which are smaller than 1. They range between 0·68 and 0·91.*

(*c*) The interferometric investigation of the light scattered by glasses shows no Doppler components (Ramm 1934; Gross 1930).

In view of these facts various authors have expressed the opinion that the observed scattering is not a true property of the glass, but they ascribe it to the presence of impurities and haphazard inner strains (Schlieren). It is, of course, possible and even quite probable that such extraneous causes do reduce the accuracy of the data. However, the series of seventeen glasses investigated by Krishnan give results of remarkable uniformity and reveal definite relationships between intensity, depolarization and the chemical composition of the glasses. It is the writer's opinion that these data eliminate any doubt that at least the predominant part of the light is scattered by the glass and not by haphazard strains or impurities.

Having discarded temperature and cluster scattering in glasses it is necessary to find a new cause for the phenomenon. I propose the following hypothesis:

The scattering is due to internal strains, but the origin of these strains is not an affair of pure chance. We assume that they are an intrinsic property of every glass and that they cannot be removed by any process of annealing. Their origin might be explained as follows: As a glass is cooled below its softening point a part of the large density fluctuations existing in the liquid state are trapped and "frozen in" in a way somewhat similar to the "frozen in" dipole orientation in the formation of electrets. As the temperature is lowered further these strains gradually increase due to temperature contractions. During this process the strain distribution is altered. It conserves its randomness, but in addition to the original strains, which were all normal in the liquid, an additional random distribution of shearing strains is created.

---

* Since in his earlier papers Krishnan used the definition $\rho_h = H_h/V_h$ the writer suspected originally that the reported values referred to this older (now abandoned) definition. The reciprocal values $\frac{2}{3}$ to $\frac{3}{4}$ would be in very good agreement with the observations. Mr Krishnan kindly assured me that his values refer to the definition $\rho_h = V_h/H_h$.

The strains represent a storage of potential energy.* Since the shearing strains are a secondary product we expect to find no equipartition of energy, but more energy is stored in the form of normal than of tangential strains.

The internal displacements due to these strains are small. To account for the light scattering one must assume that they are not more than ten times larger than those created by the temperature motion. These strains can therefore not be recorded in X-ray investigations. Due to their random character they will not manifest themselves in photoelastic observations with crossed nicols.

The light scattering due to these strains will be referred to as "strain scattering". Strain scattering does not produce any frequency modification, because the strain distribution remains practically stationary. Possible time variation occur with the velocity of diffusion and can be neglected.

THEORY OF STRAIN SCATTERING

The calculation of the light scattering by a random distribution of elastic strains is not essentially different from that for temperature scattering. One could in fact obtain the results by a slight modification of the procedure given by Leontowitsch and Mandelstamm. However, the physical significance of the results becomes much clearer, and the mathematical considerations are greatly simplified, if the problem is solved by using Brillouin's method.

The displacements $\boldsymbol{u}(x, y, z)$ which are created by the strains can be represented by a set of stationary plane waves

$$\boldsymbol{u} = \Sigma \boldsymbol{u}_k = \sum_k \boldsymbol{A}(k) \sin(\boldsymbol{k}\boldsymbol{r}).$$

The wave vectors $\boldsymbol{k}$ have the length $2\pi/\lambda$. In every direction there are three waves. One is longitudinal $(\boldsymbol{A}_L \boldsymbol{k}) = 0$, and the other two are transversal $(\boldsymbol{A}_T \boldsymbol{k}) = 0$. For a unit volume the number of waves for which the $\boldsymbol{k}$ vector is within a solid angle $d\Omega$ and for which λ is in a range $d\lambda$ is $3z(\lambda)\, d\lambda\, d\Omega$, where $z(\lambda) = 1/\lambda^4$.

---

* Their presence should influence the value of the specific heat.

The density of potential energy stored in a single longitudinal or transversal wave is

$$U_L = \tfrac{1}{4}c_{11}k^2A_L^2, \quad U_T = \tfrac{1}{4}c_{44}k^2A_T^2, \tag{7}$$

respectively.

For the case of longitudinal waves in liquids Brillouin (1922) has shown that only those waves contribute to the scattering in a given direction which have a wave vector **k** which differs very little, in magnitude and in direction, from that, $\mathbf{k}_B$, given by Bragg's law. The wave-length of this "Bragg wave" is

$$\lambda_B = \lambda_0/2n\cos\psi, \tag{8}$$

and its direction bisects the angle 2ψ between the direction of observation and the direction opposite to that of the incident beam. λ_0 is the wave-length of the incident light, n the index of refraction.

The result of Brillouin's calculation can be given in the following simple form: If the incident light is plane polarized and of unit intensity the light scattered by a unit volume to a point at the distance $R = 1$ has the intensity

$$I = \frac{\pi^2}{\lambda_0^4}\frac{U(k_B)}{c}D_n^2(k_B). \tag{9}$$

The direction of vibration of the scattered light is that of the vector $\mathbf{D}_n$. $U(k_B)$ is the density of potential energy of the Bragg wave, c is the reciprocal of the adiabatic compressibility. $\mathbf{D}_n(k_B)$ is the component normal to the direction of observation of a vector $\mathbf{D} = \epsilon\mathbf{E}_0$, where $\mathbf{E}_0$ is the electric amplitude vector of the incident light and ϵ is the amplitude of the variation of the optical dielectric constant as produced by a Bragg wave of amplitude $A = 1/k_B$.

The fundamental relation (9) is derived by calculating first the scattering by a single wave and then integrating the contributions of all waves, by taking into account the distribution law $z(\lambda)$. $z(\lambda)$ does not occur in the final result because the integration introduces a compensating factor λ_B^4. Brillouin's derivation assumes that ϵ is a small scalar quantity.

An elastic wave in a solid creates photoelastic effects, i.e. the variation of the optical dielectric constant is given by a symmetrical tensor $\epsilon_{xy}\cos(\mathbf{kr})$. A repetition of Brillouin's calculations shows that this fact alters the results only in so far as c must be replaced by c_{11} for longitudinal, and by c_{44} for transversal waves, and **D** must be defined as a vector with the components

$$\left.\begin{aligned} D_x &= \epsilon_{xx}E_{0x} + \epsilon_{xy}E_{0y} + \epsilon_{xz}E_{0z}, \\ D_y &= \epsilon_{yx}E_{0x} + \epsilon_{yy}E_{0y} + \epsilon_{yz}E_{0z}, \\ D_z &= \epsilon_{zx}E_{0x} + \epsilon_{zy}E_{0y} + \epsilon_{zz}E_{0z}. \end{aligned}\right\} \tag{10}$$

ϵ_{yz} are the components of the tensor which determines the amplitudes of that variation of the optical dielectric constant which is produced by a longitudinal or transversal Bragg wave with the amplitude $A = 1/k_B$. There are three such Bragg waves to be considered. One is longitudinal and has the energy density $U_L(k_B)$. The other two are transversal, $U = U_T(k_B)$. For one transversal wave the direction of oscillation $\mathbf{A}_t$ can be chosen parallel to the plane Π, for the other $\mathbf{A}_T$ is normal to Π. Π is the plane parallel to the directions of incidence and observation. The tensors ϵ_{yz} and hence the vectors $\mathbf{D}$ differ for the three waves. I distinguish the three cases by the subscripts L, t and T. A second subscript, h or v, indicates that the incident light is vibrating parallel or normal, respectively, to Π. Finally, to describe the polarization of the scattered light, I resolve $\mathbf{D}_n$ into two components D^H and D^V parallel and normal to Π. Hence the twelve vector components D^H_{Lv}, etc., which enter into equations (11) given below, are required.

$$\left.\begin{aligned}
V_v &= \frac{\pi^2}{\lambda_0^4}\left[\frac{U_L}{c_{11}}(D^V_{Lv})^2+\frac{U_t}{c_{44}}(D^V_{tv})^2+\frac{U_T}{c_{44}}(D^V_{Tv})^2\right],\\
H_h &= \frac{\pi^2}{\lambda_0^4}\left[\frac{U_L}{c_{11}}(D^H_{Lh})^2+\frac{U_t}{c_{44}}(D^H_{th})^2+\frac{U_T}{c_{44}}(D^H_{Th})^2\right],\\
V_h &= \frac{\pi^2}{\lambda_0^4}\left[\frac{U_L}{c_{11}}(D^V_{Lh})^2+\frac{U_t}{c_{44}}(D^V_{th})^2+\frac{U_T}{c_{44}}(D^V_{Th})^2\right],\\
H_v &= \frac{\pi^2}{\lambda_0^4}\left[\frac{U_L}{c_{11}}(D^H_{Lv})^2+\frac{U_t}{c_{44}}(D^H_{tv})^2+\frac{U_T}{c_{44}}(D^H_{Tv})^2\right].
\end{aligned}\right\} \qquad (11)$$

These equations give the four components of the scattered light as the sum of the contributions due to the three Bragg waves.

Since the wave-length of the Bragg waves is of the order of magnitude of the wave-length of light, it is justifiable to use the macroscopic laws of photoelasticity for the calculation of the tensor ϵ_{yz}. Pockel's laws of accidental birefringence in isotropic solids are given by six equations of the form

$$\left.\begin{aligned}
\epsilon_{xx} &= \epsilon_0^2[p_{11}x_x+p_{12}y_y+p_{12}z_z],\\
\epsilon_{yz} &= \epsilon_0^2 p_{44}y_z; \quad p_{44} = \tfrac{1}{2}(p_{11}-p_{12}).
\end{aligned}\right\} \qquad (12)$$

$\epsilon_0 = n^2$ is the optical dielectric constant, p_{ik} are the elasto-optical constants, and $x_x = \partial u_x/\partial x$, $y_z = \partial u_y/\partial z + \partial u_z/\partial y$ are the elastic strains.

For a longitudinal wave it is convenient to choose the x-axis in the direction of the wave vector $\mathbf{k}$. From $\mathbf{u} = u_x = A \sin kx$ we find $x_x = Ak \cos kx$, and all other components of the strain tensor vanish. For a wave of amplitude

$A = 1/k$ the amplitudes of the variation of the optical dielectric constant are therefore, according to equations (12),

$$\left.\begin{aligned} \epsilon_{xx} &= \epsilon_0^2 p_{11}, \\ \epsilon_{yy} &= \epsilon_{zz} = \epsilon_0^2 p_{12}, \\ \epsilon_{yz} &= \epsilon_{zx} = \epsilon_{xy} = 0. \end{aligned}\right\} \tag{13}$$

Hence for a longitudinal wave the principal axes of this tensor are parallel and normal to the wave vector.

For a transversal wave two of the principal axes of the ϵ_{yz} tensor are in the plane which is parallel to $\mathbf{k}$ and $\mathbf{A}$ and they form the angles ± 45° with the direction of $\mathbf{k}$. The components of the displacements along these axes are $u_x = -u_y = \frac{A}{\sqrt{2}} \sin k(x+y)/\sqrt{2}$, $u_z = 0$, whence $x_x = -y_y = \frac{1}{2} Ak \cos kr$, $z_z = y_x = x_z = z_y = 0$ and equations (12) furnish

$$\left.\begin{aligned} \epsilon_{xx} &= -\epsilon_{yy} = \epsilon_0^2 p_{44}, \\ \epsilon_{zz} &= \epsilon_{yz} = \epsilon_{zx} = \epsilon_{xy} = 0. \end{aligned}\right\} \tag{13'}$$

For the t wave the xy plane is parallel to the Π plane, for the T wave it is normal thereto.

From equations (10) therefore

$$\begin{aligned} D_{Lx} &= \epsilon_0^2 p_{11} E_{0x}^L, & D_{tx} &= \epsilon_0^2 p_{44} E_{0x}^t, & D_{Tx} &= \epsilon_0^2 p_{44} E_{0x}^T, \\ D_{Ly} &= \epsilon_0^2 p_{12} E_{0y}^L, & D_{ty} &= -\epsilon_0^2 p_{44} E_{0y}^t, & D_{Ty} &= -\epsilon_0^2 p_{44} E_{0y}^T, \\ D_{Lz} &= \epsilon_0^2 p_{12} E_{0z}^L, & D_{tz} &= 0, & D_{Tz} &= 0, \end{aligned}$$

where E_{0x}^L are the components of the electric amplitude vector in the respective system of principal axes, $E_0 = 1$. It is now a simple matter to evaluate the twelve components D_{Lv}^V for any direction of observation (see fig. 2).

(*a*) *Longitudinal wave*

If E_0 is normal to Π, $E_{0x}^L = E_{0y}^L = 0$, $E_{0z}^L = 1$. Hence $D_L = D_{Lz} = \epsilon_0^2 p_{12}$ is normal to Π and to the direction of observation, so that

$$D_{Lv}^V = \epsilon_0^2 p_{12}, \quad D_{Lv}^H = 0.$$

If E_0 is parallel to Π, $E_{0x}^L = \cos\psi$, $E_{0y}^L = \sin\psi$, $E_{0z}^L = 0$ and D_L must be parallel to Π. Hence

$$D_{Lh}^V = 0,$$

and a simple calculation furnishes

$$D^H_{Lh} = \epsilon_0^2(p_{11}\sin^2\psi - p_{12}\cos^2\psi).$$

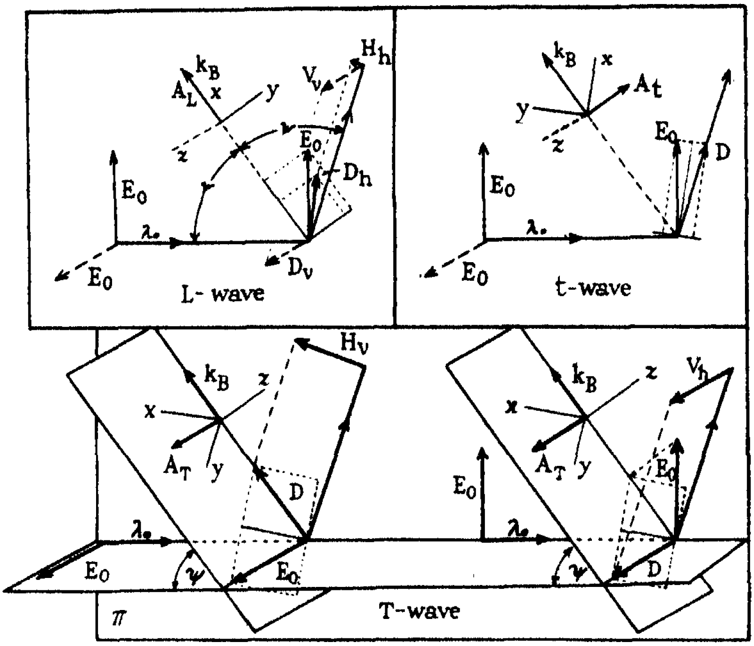


FIG. 2. Graphical determination of the vectors D which determine the scattering of light by elastic waves. In the upper figures for a longitudinal L-wave and a transversal t-wave all vectors, excepting the dotted ones, are in the plane Π. The lower figures give a perspective for a transversal T-wave. λ_0 is the direction, E_0 the amplitude of the incident light. The direction of observation is inclined at an angle 2ψ. k_B is the wave vector of the "Bragg wave", A its amplitude. x, y, z are the principal axes of the ϵ_{yz} tensor.

(b) *Transversal wave t*

If E_0 is normal to Π it follows from $\epsilon_{zz} = 0$ that $D_t = 0$, whence

$$D^V_{tv} = D^H_{tv} = 0.$$

If E_0 is parallel to Π, $E^t_{0x} = \cos(\psi + 45°)$, $E^t_{0y} = \sin(\psi + 45°)$, and, since $\epsilon_{xx} = -\epsilon_{yy}$, it can readily be shown that D_t is parallel to the direction of observation. Hence the normal component vanishes and

$$D^V_{th} = D^H_{th} = 0.$$

Hence the t waves produce no scattering. This does, however, not imply that all transversal waves are ineffective for light scattering, as has frequently been assumed. The recent investigations of the optical diffraction by supersonic waves in solids (Schaefer and Bergmann 1935; Hiedemann and Hoesch 1936; Mueller 1937) have definitely established the fact that transversal waves do create optical effects, and it will be found that the T waves give a contribution to the scattering.

(c) *Transversal waves T*

The x and y axes are at ± 45° to $\mathbf{k}_B$ in the plane normal to Π. If E_0 is normal to Π, $E^T_{0x} = -E^T_{0y} = 1/\sqrt{2}$, $E^T_{0z} = 0$ and from $D_x = D_y = \epsilon_0^2 p_{44}/\sqrt{2}$ it follows that D_T has the magnitude $\epsilon_0^2 p_{44}$ and is parallel to the plane Π. The component normal to the direction of observation gives therefore

$$D^H_{Tv} = \epsilon_0^2 p_{44} \sin\psi, \quad D^V_{Tv} = 0.$$

If E_0 is parallel to Π, $E^T_{0x} = E^T_{0y} = \sin\psi/\sqrt{2}$, hence

$$D_{Tx} = -D_{Ty} = \epsilon_0^2 p_{44} \sin\psi/\sqrt{2}, \quad D_{Tz} = 0,$$

which means that $D_T = \epsilon_0^2 p_{44} \sin\psi$ is normal to the plane Π and therefore

$$D^H_{Th} = 0, \quad D^V_{Th} = \epsilon_0^2 p_{44} \sin\psi.$$

This completes the calculation. Of the twelve components only four do not vanish. Introducing their values in equation (11) and replacing $\epsilon_0 = n^2$ the final result is

$$\left.\begin{aligned} V_v &= \frac{\pi^2 n^8}{\lambda_0^4} \frac{U_L}{c_{11}} p_{12}^2, \\ H_h &= \frac{\pi^2 n^8}{\lambda_0^4} \frac{U_L}{c_{11}} (p_{11} \sin^2\psi - p_{12} \cos^2\psi)^2, \\ H_v = V_h &= \frac{\pi^2 n^8}{\lambda_0^4} \frac{U_T}{c_{44}} p_{44}^2 \sin^2\psi. \end{aligned}\right\} \qquad (14)$$

The components V_v and H_h are due to scattering by the longitudinal Bragg wave only, i.e. they depend exclusively on the normal strains. The components H_v and V_h, however, are created by the transversal T wave, hence they originate exclusively from those shearing strains which act in planes normal to the plane of observation Π. The reciprocity relation $H_v = V_h$ is valid not only for transverse but all directions of observation. In our theory this relation arises from the fact that H_v and V_h originate from the same wave which has a particularly simple ϵ_{yz} tensor.*

For plane polarized incident light the light scattered by either Bragg wave is always plane polarized, but the planes of polarization are different for the light from the L and T waves. In the experiments it is usually not possible to separate these two light components. Hence the scattered light appears partly depolarized.

The derivation of equations (14) makes no assumptions about the origin of the strains. Due to their general validity they are applicable to various problems of light scattering.

TEMPERATURE SCATTERING IN SOLIDS

If the elastic strains are produced by temperature fluctuations the densities of potential energy are $U_L = U_T = \frac{1}{2}kT$. For transverse observation, $\psi = 45°$, and, since $\frac{1}{2}(p_{11} - p_{12}) = p_{44}$, we regain the result, equation (3) of Leontowitsch and Mandelstamm and verify the equations (5) which predict the Krishnan effect for solids.

Since the thermal elastic waves travel with the velocity of sound the scattered light suffers frequency shifts due to Doppler effect. Since the sound velocities differ for the longitudinal and transversal waves the interferometric investigations should show four Doppler components $\nu_0 \pm \nu_L$ and $\nu_0 \pm \nu_T$ for isotropic solids. If the incident light is polarized each component should also be plane polarized. The ratio of the Doppler shifts is $\frac{\nu_T}{\nu_L} = \sqrt{\frac{1-2\sigma}{(1-2\sigma)}}$ or about 0·5–0·6. The intensity ratio of the two sets of components is, for transverse observation, $\frac{I_L}{I_T} = \frac{1-2\sigma}{1-\sigma}\left[1 + \left(\frac{p_{12}}{p_{44}}\right)^2\right]$. The photoelastic measurements show that p_{12} is usually from 4 to 10 times larger than p_{44}. Hence in the most favourable case I_T is less, and usually much less, than 10% of I_L. For other directions of observation the intensity ratio is still less favourable for the experimental investigation.†

---

* Hiedemann and Hoesch (1936) have found that the light diffracted by transversal waves has very simple polarization properties. This is due to exactly the same cause which is responsible for the reciprocity relation, see Nath and Mueller (1938).

† In a recent paper E. Hiedemann (1938) points out the reasons for expecting two sets of Doppler components in solids. He suggests that the unmodified frequency component observed in liquids might be due to transversal waves. However, in liquids p_{44} must be more than 100 times smaller than p_{12}, judging from the Maxwell constant of streaming birefringence. Hence this suggestion cannot explain the observed intensity ratios. Also the observed polarization of the unmodified line is contrary to this interpretation. I am indebted to Dr Hiedemann for the privilege of reading his manuscript.

For crystals the scattered light should consist of three pairs of Doppler components because there are three elastic waves of different velocities in every direction. All three can contribute to the scattering, except for special directions of incidence and observation. If these directions are chosen properly the intensities of all six lines can be made to differ very little. Investigations of this type on crystals (Gross 1930) could produce a great deal of new information. According to Landsberg (1929, 1931) and his collaborators the light scattered from carefully selected crystals is largely due to temperature fluctuations.

STRAIN SCATTERING IN GLASSES

For glasses it is customary to replace the elasto-optical constants p_{ik} by Neumann's strain-optical constants q and p

$$p_{11} = 2q/n, \quad p_{12} = 2p/n, \quad p_{44} = (q-p)/n.$$

Since the temperature fluctuations can explain only a very small part of the observed intensities we neglect them and assume that in Krishnan's experiments one is dealing only with scattering by the hypothetical "frozen in" strains. For transverse observation ($\psi = 45°$) the equations (14) furnish

$$\rho_v = \frac{H_v}{V_v} = \frac{1}{4}\frac{U_T}{U_L}\frac{1-\sigma}{1-2\sigma}(1-q/p)^2, \tag{15}$$

$$\rho_h = \frac{V_h}{H_h} = \frac{U_T}{U_L}\frac{1-\sigma}{1-2\sigma}. \tag{16}$$

Hence

$$\frac{\rho_v}{\rho_h} = \tfrac{1}{4}(1-q/p)^2, \tag{17}$$

and the total intensity is

$$I = \frac{4\pi^2 n^6}{\lambda_0^4}\frac{U_L}{c_{11}}p^2(1+\rho_v/\rho_h+2\rho_v). \tag{18}$$

According to equation (17) the ratio ρ_v/ρ_h depends only on the ratio of the strain-optical constants and is therefore independent of the magnitude or origin of the strains. This relation provides the most important test of our theory. We give therefore in Table I Krishnan's measurements arranged according to the values of ρ_v/ρ_h. Unfortunately we do not know the photo-elastic constants of Krishnan's glasses. A fairly large amount of photo-elastic data on other glasses are available, but in most cases they give only the difference $p-q$, while we require the ratio q/p. This ratio is known only for one borate, one borosilicate and five lead glasses from the measurements of Pockels (1906). His results and the values of ρ_v/ρ_h, calculated from equation (17), are given in Table II. Comparison of the ρ_v/ρ_h columns of the two tables clearly reveals the validity of equation (17). All observed values, with the exception of two, fall within the range of the calculated ones. The two exceptions occur for two very light borosilicates with refractive indices smaller than any of Pockel's glasses. In a theory of the photo-elastic effect the writer (1935, 1938*a*) has shown that p and q depend on the refractive index and two quantities L_1 and L_2. The latter determine the change of

TABLE I. KRISHNAN'S MEASUREMENTS ON GLASSES. N IS THE INDEX NUMBER IN KRISHNAN'S PAPER

N	$\frac{\rho_v}{\rho_h}$ %	ρ_h %	I	$n-1$	Composition, over 10 % of
1	16·3	78	0·8	0·467	SiO_2, B_2O_3, Al_2O_3, K_2O
2	14·6	82	1·0	0·492	SiO_2, B_2O_3, K_2O
3	4·9	72	1·9	0·502	SiO_2, K_2O
5	4·5	80	0·9	0·527	SiO_2, B_2O_3
6	4·1	85	1·2	0·529	SiO_2, B_2O_3, Sb_2O_3
12	4·0	83	1·6	0·598	SiO_2, PbO
10	3·7	88	1·3	0·570	SiO_2, PbO
8	3·3	91	2·1	0·545	SiO_2, PbO
17	3·2	87	2·2	0·785	SiO_2, PbO
7	3·1	78	2·1	0·537	SiO_2, Na_2O, PbO
14	3·1	75	—	0·647	SiO_2, PbO
16	2·9	87	1·8	0·713	SiO_2, PbO
9	2·6	81	1·8	0·570	SiO_2, ZnO, BaO
11	2·6	80	1·6	0·589	SiO_2, B_2O_3, BaO
15	2·6	77	1·9	0·650	SiO_2, BaO, PbO
13	2·0	80	2·4	0·616	SiO_2, BaO, PbO
4	0·7	68	3·3	0·509	SiO_2, B_2O_3, ZnO

TABLE II. STRAIN SCATTERING IN GLASSES CALCULATED FROM POCKEL'S ELASTIC AND PHOTOELASTIC DATA

p	q	σ	E	$\frac{\rho_v}{\rho_h}$ %	$\frac{1-\sigma}{1-2\sigma}$	I	$n-1$	Composition
0·178	0·097	0·187	8·0	5·2	1·30	0·69	0·512	SiO_2, B_2O_3, Na_2O
0·182	0·110	0·274	4·8	3·9	1·61	0·67	0·507	B_2O_3, Al_2O
0·187	0·118	0·250	5·5	3·4	1·50	1·00	0·545	SiO_2, B_2O_3, PbO
0·195	0·135	0·222	6·1	2·4	1·40	1·00	0·570	SiO_2, Pbo (33 %)
0·205	0·160	0·224	5·5	1·2	1·41	1·00	0·644	SiO_2, PbO (52 %)
0·202	0·185	0·239	5·5	0·25	1·46	1·35	0·751	SiO_2, PbO (67 %)
0·218	0·237	0·260	5·0	0·18	1·54	2·19	0·962	SiO_2, PbO (80 %)

the optical polarizabilities of the atoms in an elastically deformed solid. They depend on the chemical composition of the glass. This theory gives

$$q/p = (1+\tfrac{4}{5}r-L_1)/(1-\tfrac{2}{5}r-L_2), \tag{19}$$

where $r = (n^2-1)/(n^2+2)$. The "optical deformabilities" have the values $L_1 = 0{\cdot}6$ to $1{\cdot}0$, $L_2 = 0{\cdot}1$ to $0{\cdot}3$. From equations (19) and (17) one estimates that for the lightest glasses ρ_v/ρ_h can reach values of about 15 %. Since the exceptions occur for the lightest glasses they must be considered as confirming the theory. On the other hand, it is noted that among the last five glasses in Table I there are some with rather small refractive indices. All these glasses contain bivalent cations Zn or Ba. From the work of Fajans and Joos it is known that the valency of the cations has a marked influence on the polarizabilities of the anions. Hence it is natural to expect that they also influence the "deformabilities". From equation (19) it is seen that a small increase of L_1 produces a large reduction of ρ_v/ρ_h, and the influence of the higher valent ions is therefore not quite unexpected. There are no photoelastic data which could clear up this point.

Equation (17) is a very general relation between the elasto-optical properties of an isotropic medium and the depolarizations of the light scattered by the substance. It must be valid also for temperature scattering. An experimental verification, based on photoelastic and depolarization measurements on the same glass, is very desirable. The optical diffraction by supersonic waves offers new methods for the determination of the strain-optical constants (Bergmann 1936; Fues 1936; Mueller 1938*b*). Since these methods furnish q/p in the most direct way they will facilitate the experimental verification.

According to equation (16) ρ_h does not depend on the photoelastic properties, but on Poisson's ratio. Since there exists no theoretical nor empirical relation between the elastic and optical properties (see Table II) we cannot expect any regularity of the ρ_h values in the arrangement of Table I. The data do not show any. Table II gives the values of $(1-\sigma)/(1-2\sigma)$ for Pockel's glasses. They are about twice as large as the observed ρ_h values. Hence equation (16) leads to the conclusion that for all glasses U_T must be considerably smaller than U_L. This agrees with our expectation that there is no equipartition of energy for these strains. The observed fact $\rho_h < 1$ must mean that the shearing strains are smaller than the normal strains.

Using equation (18) we have calculated relative values for the total scattered intensity. If U_L is assumed to be constant for all glasses it is found that I increases with increasing values of the elasto-optical constant p. Consequently glasses with a large value of ρ_v/ρ_h can be expected to scatter less light than similar glasses with a small ρ_v/ρ_h. Krishnan's intensity data confirm this result. All glasses with a small scattering power appear in the upper part of Table I, and all large values of I are associated with small values of ρ_v/ρ_h.

These striking analogies between the calculated and the observed data are convincing enough to accept the interpretation of scattering by a random distribution of strains. They eliminate any doubt that one is dealing with a true property of the glass.

STRUCTURE SCATTERING

From the point of view of the continuum theory of light scattering there is no difference between a medium with internal strains and a substance with molecular clusters. *We might indeed say that the strains provide the glass* with properties which make it appear as if it contained clusters. It is on account of Warren's work that the clusters in glasses must be considered as fictitious. In mixtures of liquids near the critical point, however, the hypothesis of cluster formation is generally accepted. A "clustered" liquid can be considered as if it were a uniform medium in which internal strains, in this case due to intramolecular forces, have created local density variations and local optical anisotropy. From the point of view of a molecular theory the origin of the Krishnan effect is probably not the same for glasses and for liquids, but since the wave-length of light is large compared to molecular dimensions the ordinary measurements on scattered light (excluding spectroscopic investigations) do not permit to distinguish between the two cases.

The essential assumption for the explanation of the Krishnan effect is not the existence of internal strains. The strains have been introduced only to evaluate the variations of the optical dielectric constant. Essential are the assumptions that these variations are described by a tensor η_{xy}, and that these fluctuations are not independent for neighbouring volume elements. The second assumption leads to the representation by means of Fourier series or by a set of plane waves

$$\eta_{yz} = \sum_{k} e_{yz}(k)\cos(\boldsymbol{kr} + \phi_{yz}). \tag{20}$$

The phase angles $\phi^{(k)}_{yz}(k)$ must be introduced to take into account that the four components V_v, H_h, H_v and V_h need not be coherent. If the x axis points in the direction of **k** the isotropy of the medium requires that the tensor e_{yz} be of the form

$$|e_{yz}| = \begin{vmatrix} a & c & c \\ c & b & 0 \\ c & 0 & b \end{vmatrix}. \tag{21}$$

By adding the three tensors for the L, t and T wave, we see indeed that this representation is equivalent to that for strain scattering. All previous results hold if one substitutes

$$\left.\begin{aligned} a(k) &= A_L(k)\, k\epsilon_0^2\, p_{11}, \\ b(k) &= A_L(k)\, k\epsilon_0^2\, p_{12}, \\ c(k) &= A_T(k)\, k\epsilon_0^2\, p_{44}. \end{aligned}\right\} \tag{22}$$

Taking into account equation (7) our results, equations (14), can therefore be given in the more general form

$$\left.\begin{aligned} V_v &= \tfrac{1}{4}\pi^2 \frac{1}{\lambda_0^4} b^2, \\ H_h &= \tfrac{1}{4}\pi^2 \frac{1}{\lambda_0^4} (a\sin^2\psi - b\cos^2\psi)^2, \\ H_v = V_h &= \tfrac{1}{4}\pi^2 \frac{1}{\lambda_0^4} c^2 \sin^2\psi, \end{aligned}\right\} \tag{23}$$

and the depolarizations for transverse observation are

$$\begin{aligned} \rho_v &= \tfrac{1}{2}c^2/b^2, \\ \rho_h &= 2c^2/(a-b)^2, \end{aligned} \qquad \rho_v/\rho_h = \tfrac{1}{4}(1-a/b)^2, \tag{24}$$

where a, b and c are the coefficients of the Fourier series (20) for

$$k = 2\pi/\lambda_B = 4\pi n \cos\psi/\lambda_0.$$

Equations (23) and (24) are valid for any isotropic medium in which the fluctuations of the optical anisotropy can be represented by a Fourier series. There can be little doubt that this is the case for liquids with molecular clusters. Within certain limitations this theory can also be applied to colloids. It can, of course, not be used for sols with metallic particles, but if the micelles are not much larger than the wave-length of light and consist of a transparent substance, a Fourier series seems an adequate representation of the variation of the refractive properties. It appears therefore likely that the theory may be valuable also for the study of certain colloids, proteins and resins.

The equations (22) correlate the Fourier coefficients of the optical fluctuations to the internal strains and to the strain-optical constants. They are analogous to the Gans relation between the depolarization and the Kerr constant. While in Gans' derivation the Kerr effect serves to evaluate the energy which is required to produce optical anisotropy, in our theory this energy is given by the elastic strains which can create the same effect. The data in Tables I and II are a verification of these relations for glasses. For viscous liquids a similar test ought to be possible if p_{11}, p_{12} and p_{44} are interpreted as the constants of streaming birefringence. The usual methods of measuring streaming birefringence employ shearing strains and hence furnish only p_{44}. However, the recently discovered method for measuring p_{11}/p_{12} in glasses by optical diffraction on supersonic longitudinal waves can probably also be used for determining this ratio for viscous liquids. A test of the relations requires only the knowledge of p_{11}/p_{12}, because it is sufficient to show that for transverse light scattering $\rho_v/\rho_h = \frac{1}{4}(1-p_{11}/p_{12})^2$ (equation 17).

From the observed fact $\rho_h < 1$ it follows, according to equation (24), that in all liquids $\frac{1}{2}(a-b) > c/\sqrt{2}$. This inequality can be understood by interpreting the optical fluctuations as due to internal strains. Equations (22) give then $A_L > A_T/\sqrt{2}$, i.e. it means that the normal strains are considerably larger than the shearing strains. As this should naturally be expected for any liquid, it is not surprising that the case $\rho_h > 1$ has never been observed.

Measurements of the depolarization ρ_h can give information about the temperature motion in liquids. A more rigorous theoretical interpretation should take into account that the longitudinal and the transversal waves are damped (Lucas 1937) and that a part of the fluctuations cannot be represented by waves.

Equations (23) verify again the validity of Krishnan's reciprocity relation $V_h = H_v$. The necessary and sufficient condition for obtaining this law is that e_{yz} is a symmetrical tensor. Hence it holds for any liquid or amorphous solid, excepting those which are optically active.*

STRUCTURE ANALYSIS BY MEANS OF LIGHT SCATTERING

Gans' theory of the Krishnan effect makes the assumption that the molecular clusters are of uniform size and have a definite shape. In our representation the density and optical anisotropy vary continuously. Our theory can give no information about the shape of the clusters, but it opens a new way for determining their average size.

---

* Krishnan (1938) attempts to prove that H_v/V_h is a true reciprocity relation, i.e. that it can be derived by exchanging the role of source and observer. However, his proof is not rigorous because it is based on principles of acoustics and not of electro-magnetic theory.

By using light of different wave-lengths and by varying the direction of observation the wave-length $\lambda_B = \lambda_0/2n\cos\psi$ of the Bragg wave can be altered continuously from about 30,000 to 1500 A. Hence, from measurements of the intensity of the three components V_v, H_h and V_h or H_v as functions of λ_0 and ψ the functions $a(\lambda_B)$, $b(\lambda_B)$ and $c(\lambda_B)$ can be determined for a large range of λ_B. For temperature scattering in solids no large variations of a, b and c can be expected because we have equipartition of energy and the dispersion of the elasto-optical constants is small. For strain- and structure-scattering, however, these functions may show large variations. If the fluctuations would follow the law of Ornstein and Zernicke a, b and c should, at the critical point, be proportional to λ_B and we would obtain their λ_0^{-2} law for the intensity of the scattered light. Since most observations do not agree with this law, further progress in the problem of opalescence can only be made by reversing the procedure, i.e. instead of postulating the type of fluctuations, one should use the measurements for finding out the nature of the phenomenon. The data furnish a Fourier analysis of the optical variations and of the internal strains. From it we can gain information about the size of the clusters or of the colloidal particles. If, for instance, it should be found that a, b and c have maxima for $\lambda_B = d$, it would be natural to associate this length with the average size of the clusters.

This proposed structure analysis is very similar to the method of Zernicke and Prins (1927) for the X-ray analysis of liquid structures. It is obvious that scattering of visible light can furnish information about the coarser structure of matter by using the same mathematical tools which enable one to find the molecular structure of liquids and amorphous solids from the scattering of X-rays.

λ_0^{-m} LAW AND MIE EFFECT

The dependence of the intensity of the transversely scattered light on the wave-length has been the subject of a number of investigations (Andant 1922; Bhatta Charyya 1923; Rocard 1928; Rousset 1935). The critical opalescence of some liquids follows the λ_0^{-4} law, but for many mixtures a law λ_0^{-m} with $m < 4$ has been observed. At the critical point the depolarizations do not vanish, but they are so small that over 95 % of the total intensity is due to the component V_v.

On the basis of our theory the λ_0^{-4} is satisfied if $b(\lambda_B)$ is constant for wave-lengths within the visible spectrum, i.e. if the clusters are much smaller or much larger than the wave-length of light, or if there are no clusters. $m < 4$ implies that $b(\lambda_B)$ is an increasing function. Hence the clusters must be larger than $\lambda_0/n\sqrt{2}$. The case $m > 4$ is not excluded by our theory. It could occur for a liquid with very small clusters.

For observations in an arbitrary direction the components V_v and H_h give the major contribution to the total intensity. In liquids a and b differ by very little (small streaming birefringence). Hence, in first approximation, the total intensity for observations in a direction 2ψ is

$$I(\psi) = \tfrac{1}{4}\pi^2 \frac{b^2}{\lambda_0^4}(1+\cos^2 2\psi). \tag{25}$$

If $b(\lambda_B)$ is constant, i.e. for mixtures which satisfy the λ_0^{-4} law, the intensity distribution is symmetrical with respect to the direction normal to the incident beam, $I(45^\circ+\alpha) = I(45^\circ-\alpha)$. For liquids with $m < 4$ we must have $b^2 \sim \lambda_B^{4-m}$ and obtain, since $\lambda_B = \lambda_0/2n\cos\psi$,

$$M(\alpha) = \frac{I(45^\circ+\alpha)}{I(45^\circ-\alpha)} = \left[\frac{\cos(45^\circ-\alpha)}{\cos(45^\circ+\alpha)}\right]^{4-m}. \tag{26}$$

Hence if $m \neq 4$ the intensity distribution is asymmetrical. For $m < 4$ more light is scattered forward, in the direction of propagation of the incident light, than backward. This effect is called the Mie effect. Liquids with $m > 4$ should show a reversed or "negative" Mie effect.

The conclusion that clustered liquids show a Mie effect if they do not satisfy the λ_0^{-4} law, is in excellent agreement with the observations of Rousset (1935). For three critical mixtures Rousset verified the λ_0^{-4} law and could observe no Mie effect. For the mixtures water-isobutyric acid and aniline-cyclohexane at their critical points he finds $m = 2{\cdot}9$ and $3{\cdot}6$, respectively. Rousset also measured the intensity ratios M for two symmetrical directions at 37° to the normal and found for the first mixture $M = 2{\cdot}24$ and for the second $M = 1{\cdot}66$. For $\alpha = 18{\cdot}5^\circ$ equation (26) would predict $M = 2{\cdot}15$ and $1{\cdot}32$, respectively. The agreement is surprisingly good, considering the approximation involved. Since no other theory has been able to account for Rousset's observations this result is a very strong point in favour of our theory.

By using polarized incident light and observing the scattering through a nicol it is possible to study the Mie effect of each component V_v, H_h and $H_v = V_h$. V_v will have a Mie effect only if the λ_0^{-4} does not hold. H_v must always show a positive Mie effect. The Mie effect of H_h can be positive or negative. It is negative if the deviations from the λ_0^{-4} law are small and if $a < b$. This case is of particular interest because it reveals a connexion between Mie effect and streaming birefringence. Equations (22) show that $a < b$ when $p_{11} - p_{12} = 2p_{44} < 0$. This is the case for most glasses ($q - p < 0$, see Table II) and for liquids with positive streaming birefringence. Hence it appears very probable that glasses and semi-solid substances with a large positive streaming birefringence should show an inversed Mie effect. This is perhaps the explanation for the "scattering reflexion" observed in resins (Coban 1935).

All substances which have a Krishnan effect should show a Mie effect, though the latter will be very small if the depolarization ρ_v is small. Since for critical opalescence ρ_v is usually less than 1 % one can understand that Rousset could not find a Mie effect in the mixtures water-triethylamine and nitrobenzene-hexane despite the fact that these mixtures have a Krishnan effect.

If the dependence on λ_B is different for $a(\lambda_B)$, $b(\lambda_B)$ and $c(\lambda_B)$ the depolarizations must have a dispersion (Krishnan 1936*c*) and the angular intensity distribution can become very asymmetrical. The equations (23) could offer an explanation of the disputed "Plotnikov effect" for small angle longitudinal scattering (Coban 1935, Katalinić 1937).

This theory applies only to that part of the scattering which is due to fluctuations which can be represented by a Fourier series. In ordinary liquids not all fluctuations satisfy this condition, as is shown by the appearance of the unmodified frequency component. The recent results of Raman and Rao (1938), which show that the Doppler components vanish for liquids with a high viscosity, are perhaps an indication that structure scattering is the predominant process in viscous media.

SUMMARY

The Krishnan effect can be explained if the fluctuations of the optical anisotropy in the medium are not independent in neighbouring volume elements. By applying Brillouin's method to longitudinal and transversal waves the Krishnan effect is calculated for arbitrary directions of observation. It is found that the reciprocity relation is always valid. Every substance which has a Krishnan effect should show a Mie effect. Both effects are related to the photoelastic properties in the case of solids, and to the constants of streaming birefringence in the case of liquids.

The Krishnan and Mie effects must occur for temperature scattering in solids. In isotropic solids the scattered light must consist of two pairs of Doppler components, for crystals three different Doppler shifts can be expected. The Krishnan effect can occur in ordinary liquids if Lucas'

transversal waves give a noticeable contribution to the scattering. For temperature scattering the depolarization $\rho_h = V_h/H_h$ is always larger than 1.

A new theory of the scattering of light in glasses is proposed. It is based on the assumption that glasses contain a random distribution of "frozen in" strains. Slightly below the solidification temperature these strains are normal ones, but at lower temperatures shearing strains are created due to temperature contraction. ρ_h is smaller than 1 because the shearing strains are always smaller than the normal strains. It is found that Krishnan's data are in excellent agreement with those calculated for glasses for which the photoelastic constants are known.

Liquids with molecular clusters can be treated as if they were uniform liquids with a distribution of internal strains. They show $\rho_h < 1$ because the strains are predominantly normal ones. A relation between the Mie effect for critical opalescence and the deviation from the λ^{-4} law is found to agree with Rousset's observation.

It is pointed out that scattering data furnish a Fourier analysis of the optical variations within the medium and can be used to determine the size of molecular clusters and of colloidal particles.

Gans' theory of scattering by molecular clusters is discussed and it is shown that it agrees with Krishnan's reciprocity relation.

REFERENCES

Andant, A. 1922 *C.R. Acad. Sci., Paris*, **174**, 1333, 1541.
Bergmann, L. 1936 *Naturwissenschaften*, **24**, 492.
Bhatta Charyya, D. K. 1923 *Proc. Indian Ass. Cult. Sci.* **8**, 277.
Brillouin, L. 1922 *Ann. Phys., Paris*, **17**, 88–122.
— 1936 *J. Phys. Radium*, **7**, 153–7.
Coban, B. 1935 *Acta phys. polon.* **4**, 1–16.
Fues, E. 1936 *Naturwissenschaften, Lpz.*, **24**, 492.
Gans, R. 1925 *Ann. Phys., Lpz.*, **77**, 317–24.
— 1936 *Phys. Z.* **37**, 19–22.
— 1937 *Phys. Z.* **38**, 625.
Gehmann, S. D. and Field, J. E. 1937 *Industr. Engng. Chem.* **29**, 793–9.
Gross, E. 1930 *Z. Phys.* **63**, 685–7.
Hiedemann, E. 1938 *Z. Physik.* **108**, 592.
Hiedemann, E. and Hoesch, K. H. 1936 *Naturwissenschaften*, **24**, 60.
Katalinić, M. 1937 *Z. Phys.* **106**, 439–52.
Krishnan, R. S. 1934*a* *Proc. Ind. Acad. Sci.* **1**, 211–16.
— 1934*b* *Proc. Ind. Acad. Sci.* **1**, 717–22.
— 1934c *Proc. Ind. Acad. Sci.* **1**, 782–8.
— 1934*d* *Proc. Ind. Acad. Sci.* **1**, 915–27.
— 1935 *Proc. Ind. Acad. Sci.* **2**, 221–31.
— 1936*a* *Proc. Ind. Acad. Sci.* **3**, 126–34.
— 1936*b* *Proc. Ind. Acad. Sci.* **3**, 211–20.

Krishnan, R. S. 1936*c* *Proc. Ind. Acad. Sci.* **3**, 566–71.
— 1937*a* *Proc. Ind. Acad. Sci.* **5**, 94.
— 1937*b* *Proc. Ind. Acad. Sci.* **5**, 305.
— 1937*c* *Proc. Ind. Acad. Sci.* **5**, 407.
— 1937*d* *Proc. Ind. Acad. Sci.* **5**, 499.
— 1937*e* *Proc. Ind. Acad. Sci.* **5**, 551.
— 1937*f* *Proc. Ind. Acad. Sci.* **5**, 577.
— 1937*g* *Curr. Sci.* **6**, 90.
— 1938 *Proc. Ind. Acad. Sci.* **7**, 21.
Landsberg, Gr. and Mandelstamm, S. jr. 1931 *Z. Phys.* **73**, 502.
Landsberg, Gr. and Wulfson, K. 1929 *Z. Phys.* **58**, 95.
Leontowitsch, M. and Mandelstamm, S. jr. 1932 *Phys. Z. Sowjet*, **1**, 317.
Lucas, R. 1937 *J. Phys. Radium*, **8**, 410.
Mie, G. 1908 *Ann. Phys., Lpz.*, **25**, 377–445.
Mueller, H. 1935 *Physics*, **6**, 179.
— 1937 *Phys. Rev.* **52**, 223.
— 1938*a* *J. Amer. Ceram. Soc.* **21**, 27.
— 1938*b* *Z. Kristallogr.* **99** (in the press).
Nath, N. S. N. and Mueller, H. 1938 *Nature, Lond.*, **138**, 37.
Ornstein, L. S. and Zernicke, F. 1926 *Phys. Z.* **27**, 761.
Ornstein, L. S. and Van Cittert, P. H. 1935 *Physica*, **2**, 499.
Placzek, G. 1930 *Phys. Z.* **31**, 1052.
Pockels, F. 1906 "Lehrbuch der Kristallphysik." Berlin and Leipzig.
Ramaiah, K. S. 1937 *Proc. Ind. Acad. Sci.* **5**, 128, 138.
Raman, C. V. and Rao, B. V. R. 1938 *Nature, Lond.*, **141**, 242.
Ramm, W. 1934 *Phys. Z.* **35**, 756.
Rocard, Y. 1928 *Ann. Phys., Paris*, **10**, 116–231.
Rousset, A. 1935 *C.R. Acad. Sci., Paris*, **199**, 716.
Schaefer, Cl. and Bergmann, L. 1935 *S.B. preuss. Akad. Wiss.* **14**, 222.
Schaefer, Cl., Bergmann, L. and Goehlich, H. J. 1937 *Glastech. Ber.* **15**, 447–61.
Subbaramaija, D. S. 1934 *Proc. Ind. Acad. Sci.* **1**, 709–16.
— 1935 *Proc. Ind. Acad. Sci.* **2**, 358–63.
Vrkljan, V. S. and Katalinić, M. 1936 *Phys. Z.* **37**, 482–7.
Warren, B. E. 1937 *J. Appl. Phys.* **8**, 645.
Zernicke, F. and Prins, J. A. 1927 *Z. Phys.* **41**, 184–94.

Krishnan, R. S. 1936c *Proc. Ind. Acad. Sci.* **3**, 566–71.
— 1937a *Proc. Ind. Acad. Sci.* **5**, 94.
— 1937b *Proc. Ind. Acad. Sci.* **5**, 305.
— 1937c *Proc. Ind. Acad. Sci.* **5**, 407.
— 1937d *Proc. Ind. Acad. Sci.* **5**, 499.
— 1937e *Proc. Ind. Acad. Sci.* **5**, 551.
— 1937f *Proc. Ind. Acad. Sci.* **5**, 577.
— 1937g *Curr. Sci.* **6**, 90.
— 1938 *Proc. Ind. Acad. Sci.* **7**, 21.
Landsberg, Gr. and Mandelstamm, S. jr. 1931 *Z. Phys.* **73**, 502.
Landsberg, Gr. and Wulfson, K. 1929 *Z. Phys.* **58**, 95.
Leontowitsch, M. and Mandelstamm, S. jr. 1932 *Phys. Z. Sowjet*, **1**, 317.
Lucas, R. 1937 *J. Phys. Radium*, **8**, 410.
Mie, G. 1908 *Ann. Phys., Lpz.*, **25**, 377–445.
Mueller, H. 1935 *Physics*, **6**, 179.
— 1937 *Phys. Rev.* **52**, 223.
— 1938a *J. Amer. Ceram. Soc.* **21**, 27.
— 1938b *Z. Kristallogr.* **99** (in the press).
Nath, N. S. N. and Mueller, H. 1938 *Nature, Lond.*, **138**, 37.
Ornstein, L. S. and Zernicke, F. 1926 *Phys. Z.* **27**, 761.
Ornstein, L. S. and Van Cittert, P. H. 1935 *Physica*, **2**, 499.
Placzek, G. 1930 *Phys. Z.* **31**, 1052.
Pockels, F. 1906 "Lehrbuch der Kristallphysik." Berlin and Leipzig.
Ramaiah, K. S. 1937 *Proc. Ind. Acad. Sci.* **5**, 128, 138.
Raman, C. V. and Rao, B. V. R. 1938 *Nature, Lond.*, **141**, 242.
Ramm, W. 1934 *Phys. Z.* **35**, 756.
Rocard, Y. 1928 *Ann. Phys., Paris*, **10**, 116–231.
Rousset, A. 1935 *C.R. Acad. Sci., Paris*, **199**, 716.
Schaefer, Cl. and Bergmann, L. 1935 *S.B. preuss. Akad. Wiss.* **14**, 222.
Schaefer, Cl., Bergmann, L. and Goehlich, H. J. 1937 *Glastech. Ber.* **15**, 447–61.
Subbaramaija, D. S. 1934 *Proc. Ind. Acad. Sci.* **1**, 709–16.
— 1935 *Proc. Ind. Acad. Sci.* **2**, 358–63.
Vrkljan, V. S. and Katalinić, M. 1936 *Phys. Z.* **37**, 482–7.
Warren, B. E. 1937 *J. Appl. Phys.* **8**, 645.
Zernicke, F. and Prins, J. A. 1927 *Z. Phys.* **41**, 184–94.

Self-consistent field with exchange for potassium and argon

By D. R. Hartree, F.R.S., and W. Hartree

(*Received* 16 *March* 1938)

1. Introduction

As already mentioned in a previous paper on the calculation of the self-consistent field with exchange for calcium (D. R. and W. Hartree 1938), wave functions more accurate than those calculated without exchange are required both for K^+ and for Ar, and the calculations for calcium were carried out partly in the hope that it would be possible to interpolate with sufficient accuracy the difference between the wave functions calculated with and without exchange for K^+ and Ar from the corresponding differences for Cl^- and Ca^{++}.

The results showed that for the $(1s)$, $(2s)$ and $(2p)$ wave function this interpolation would probably be satisfactory, but for the $(3s)$ and $(3p)$ wave functions it was not as straightforward as had been hoped, though even for the latter wave functions, estimates of the differences could be made which, while rather uncertain, would probably give appreciably better wave functions than those calculated without exchange and taken without modification.

In the solution of Fock's equations (Fock 1930) for the self-consistent field with exchange, a good deal of time and labour are usually spent in finding a fairly good approximation to the solution; when once such an approximation has been obtained, the process of improving it is often not such a very long and laborious one. And since an interpolation process, applied to the differences between the wave functions calculated with and without exchange, promised to provide already good approximations to three of the wave functions and fairly good approximations to the other two, for each atom, it seemed worth while to carry out a partial solution of Fock's equations using the wave functions so found as the first estimates. For K^+, solutions of Fock's equations have been evaluated for all the wave functions except $(1s)$. For $(1s)$ the effects of exchange terms are so small that the interpolation process is certainly reliable to the accuracy required; the solutions for $(2s)$ and $(2p)$ were evaluated to test the reliability of the interpolation process as applied to them. The results of this test were satisfactory, and for Ar only the (3*s*) and (3*p*) wave functions were evaluated by actual solution of Fock's equations, the other wave functions being determined by the interpolation process.

In previous calculations of the solutions of Fock's equations, except for Ca^+ and neutral Ca, we have started with estimates of the $Z_k(nl, n'l')$ functions, following the practice, in the calculation of self-consistent field without exchange, of starting with estimates of the $Z_0(nl, nl)$ functions. But since, when exchange terms are included, the number of $Z_k(nl, n'l')$ functions is much greater than the number of radial wave functions $P(nl)$, it seems more economical to start from estimates of the latter functions. This point arose in an extreme form in the course of the calculations for Ca^+, since for each state of the series electron only one $P(nl)$ function is involved, whereas there are six $Z_k(nl, n'l')$ functions expressing the interaction of the series electron with the six $(n'l')$ groups of the core. This experience has suggested the wider use of estimates of radial wave functions rather than of Z_k functions, and in the calculations here reviewed we have worked from estimates of radial wave functions throughout.

2. THE PROCESS OF INTERPOLATION

Calculations of the self-consistent field without exchange have been carried out for K^+ (Hartree 1934) and for Ar (McDougall unpublished*), so that in order to obtain a set of initial estimates for starting the calculations with exchange, it is only necessary to interpolate the differences between the wave functions calculated with and without exchange. In most cases this is likely to be both easier and more accurate than interpolating the wave functions themselves, since the maximum value of the difference is usually of the order of a tenth or less of the maximum value of the wave function.

If $P(nl\,|\,r)$ and $\bar{P}(nl\,|\,r)$ are the normalized radial wave functions, calculated with and without exchange respectively, we will write

$$\Delta P(nl\,|\,r) = P(nl\,|\,r) - \bar{P}(nl\,|\,r); \tag{1}$$

also we will write r_m for the radius of the principal maximum of $|\,\bar{P}(nl\,|\,r)\,|$.

Since, for each (nl), the quantity to be interpolated is not just a single value of some function of the atomic number N only, but a function of r of which perhaps thirty or fifty values are required, an interpolation process which is simple and easy to use is needed, and for this purpose we want to arrange the data in such a way that interpolation is carried out on quantities which vary slowly and nearly linearly with the atomic number N.

---

* We are indebted to Dr J. McDougall, of Leeds University, for supplying us with the results of his calculations for argon.

If, for each (nl), the variation of ΔP with r were such that two constants A, B could be found for each atom so that $A\,\Delta P(nl\,|\,r)$ were the same function of Br for all atoms, the graph of this function would provide a very simple means of carrying out the interpolation.

For a given configuration, this is in fact possible in the limit $N \to \infty$ when the interaction (both Coulomb and exchange) between the electrons becomes small compared with the Coulomb interaction with the nucleus. To show this, it will be sufficient to take Fock's equation for the $(2s)$ wave function in the $(1s)^2\,(2s)^2$ configuration of a four-electron (Be-like) system. This equation is (cf. D. R. and W. Hartree 1935)

$$\left[\frac{d^2}{dr^2}+\frac{1}{r}\{2N-4Y_0(1s,1s\,|\,r)-2Y_0(2s,2s\,|\,r)\}-\epsilon_{2s,2s}\right]P(2s\,|\,r) + \frac{2}{r}Y_0(1s,2s\,|\,r)P(1s\,|\,r)=0. \tag{2}$$

Since the configuration is one of closed groups, no non-diagonal Lagrange multiplier $\epsilon_{1s,2s}$ is required; the wave functions $P(1s\,|\,r)$ and $P(2s\,|\,r)$ are supposed to be normalized.

Writing
$$Nr=\sigma \tag{3}$$
and dividing by N^2, (2) becomes

$$\left[\frac{d^2}{d\sigma^2}+\frac{1}{\sigma}\left\{2-\frac{4Y_0(1s,1s\,|\,\sigma/N)+2Y_0(2s,2s\,|\,\sigma/N)}{N}\right\}-\frac{\epsilon_{2s,2s}}{N^2}\right]P(2s\,|\,\sigma/N) + \frac{2}{N\sigma}Y_0(1s,2s\,|\,\sigma/N)\,P(1s\,|\,\sigma/N)=0. \tag{4}$$

Now let $P_{(\mathrm{H})}(nl\,|\,\sigma)$ be the hydrogen-like wave function in the field of a point charge N, normalized to $\int[P_{(\mathrm{H})}(nl\,|\,\sigma)]^2\,d\sigma=1$; this, as a function of σ, is independent of N. Also let

$$Y_{(\mathrm{H})k}(nl,n'l'\,|\,\sigma)=\int_0^\infty P_{(\mathrm{H})}(nl\,|\,\sigma_1)\,P_{(\mathrm{H})}(n'l'\,|\,\sigma_1)f_k(\sigma,\sigma_1)\,d\sigma_1,$$

where
$$\begin{aligned} f_k(\sigma,\sigma_1) &= (\sigma_1/\sigma)^k \quad \text{if} \quad \sigma_1<\sigma,\\ &= (\sigma/\sigma_1)^{k+1} \quad \text{if} \quad \sigma<\sigma_1. \end{aligned}$$

Then, to first order in $1/N$, equation (4) and the similar equation for $P(1s \mid \sigma/N)$ are satisfied by

$$P(nl \mid \sigma/N) = N^{\frac{1}{2}}\left[P_{(H)}(nl \mid \sigma)+\frac{1}{N}\,Q(nl \mid \sigma)\right], \tag{5}$$

where $Q(nl \mid \sigma)$, regarded as a function of σ, is independent of N; $Q(2s \mid \sigma)$, for example, satisfies the equation

$$\left[\frac{d^2}{d\sigma^2}+\frac{2}{\sigma}-\frac{1}{4}\right]Q(2s \mid \sigma) = \left[\frac{1}{\sigma}\{4Y_{(H)0}(1s,1s \mid \sigma)+2Y_{(H)0}(2s,2s \mid \sigma)\} + \frac{1}{N^2}\epsilon_{2s,2s}-\frac{1}{4}\right]P_{(H)}(2s \mid \sigma)-\frac{2}{\sigma}Y_{(H)0}(1s,2s \mid \sigma)\,P_{(H)}(1s \mid \sigma)\,P_{(H)}(1s \mid \sigma), \tag{6}$$

and this equation does not involve N explicitly. The factor $N^{\frac{1}{2}}$ in (5) is a normalizing factor.

The form of equation (6) shows that Q is the sum of contributions from the separate terms on the right-hand side of this equation, hence each of these contributions is a function of σ only. Now $\Delta P(2s \mid r)$ is just the contribution to $P(2s \mid r)$ from the exchange term, and from (5) this is $N^{-\frac{1}{2}} \times$ [contribution to $Q(2s \mid Nr)$ from exchange term]; hence, for large N, $N^{\frac{1}{2}}\Delta P(2s \mid r)$, regarded as a function of $\sigma = Nr$, is independent of N. A similar argument applies to the contribution to Q from the other exchange terms in configurations containing more electrons.

For finite N, a similar result would be expected to hold approximately, with N replaced by some screened value, and this suggests that for each (nl) it should be possible to find constants s_1 and s_2 such that the graph of $(N-s_1)^{\frac{1}{2}}\Delta P(nl \mid r)$ against $(N-s_2)\,r$ is approximately the same for all N. Also $Nr_m \to$ const. as $N \to \infty$; hence for finite N, $(N-s_3)\,r_m$ would be expected to be approximately constant, and if, as seems reasonable, s_2 and s_3 are not far different, the graph of $(N-s_1)^{\frac{1}{2}}\Delta P(nl \mid r)$ against r/r_m would also be nearly the same for all N, and so suitable for interpolation.

Actually since ΔP is small, and the range of N over which interpolation was to be carried out was also small, it seemed likely to be almost equally good to plot $\Delta P(nl \mid r)$ against r/r_m, and this is the form in which the interpolation was actually carried out. The curves of $\Delta P(nl \mid r)$ against r/r_m, both for Ca^{++} and Cl^- which formed the original data at the start of this work, and for K^+ and Ar calculated from the results of it, are shown in fig. 1 (*a*)–(*d*), and show the extent to which the results for the different atoms are brought into agreement by plotting in this way.

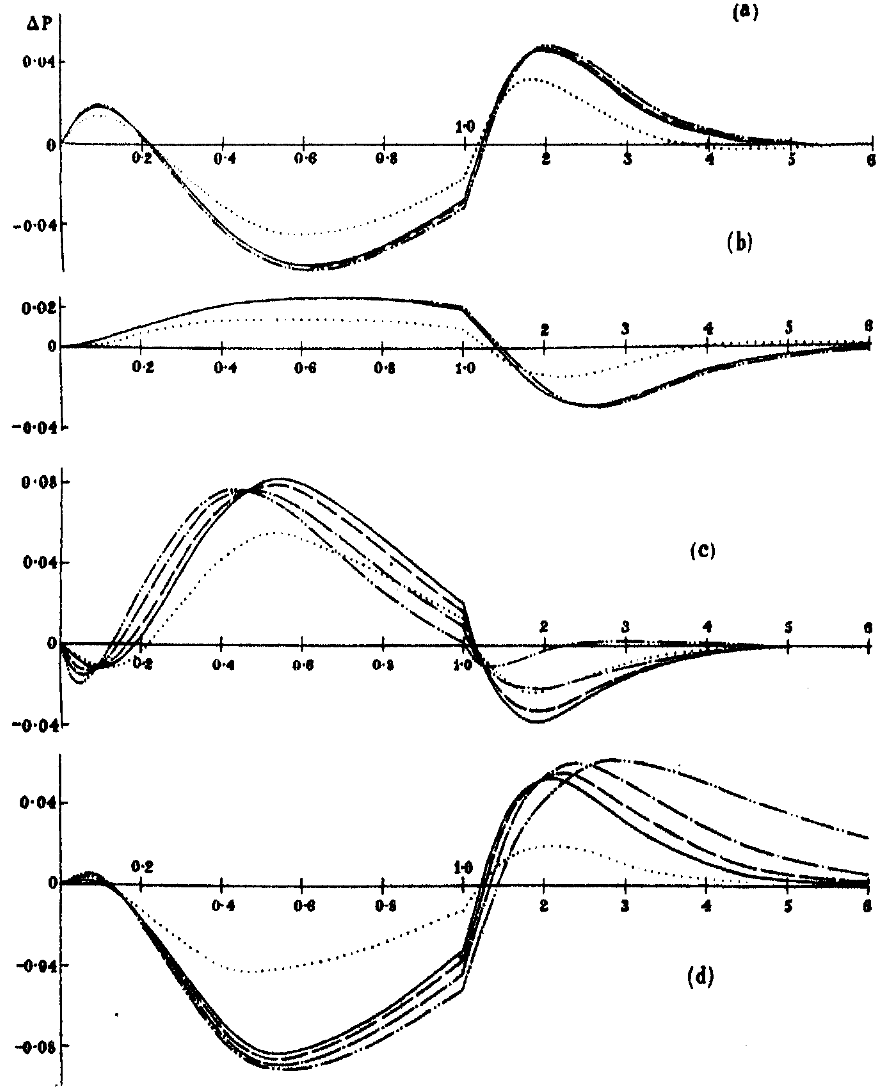


FIG. 1. Differences ΔP between normalized radial wave functions calculated with and without exchange, plotted against r/r_m, where r_m is the radius of the principal maximum of $|P|$ calculated without exchange. The results for different (nl) wave functions are shown in separate figures as follows:

Fig.	1*a*	1*b*	1c	1*d*
(nl)	$(2s)$	$(2p)$	$(3s)$	$(3p)$

and in each figure the results for different atoms are indicated as follows:

········ Cu$^+$	— — — K$^+$	—··—··— Cl$^-$
———— Ca^{++}	—·—·— Ar	

For $(2p)$, and in part of the range for $(2s)$, the curve for K$^+$ cannot be shown clearly distinct on the scale of the figures, and has been omitted. Note the change of scale of r/r_m at $r/r_m = 1$.

3. PROCESS OF SOLUTION OF FOCK'S EQUATION

The process of solution of Fock's equations was similar to that previously described (D. R. and W. Hartree 1935, 1936*a*, *b*, *c*), except that the present work started from estimates of the radial wave functions $P(nl \mid r)$ instead of from estimates of the $Z_k(nl, n'l' \mid r)$ functions, as already explained.

For K^+, the interpolation of ΔP for (1*s*), (2*s*), and (2*p*) was made assuming that the departures from a linear variation with N were negligible; that is to say, at each value of r/r_m, ΔP for K^+ was taken as one-third of the way from the value for Ca^{++} towards that for Cl^-. For (3*s*), the fact that $|\Delta P|$ for Cl^- is less than for Ca^{++} over a large part of the range, whereas by the argument of § 2 it would be expected to be greater, suggested that the departures from a linear variation with N are considerable. It was not even certain whether $|\Delta P|$ would continue to increase, with decreasing N, beyond Ca^{++} to K^+ or perhaps further, before decreasing to Cl^-, or whether the decrease had already begun at Ca^{++}; a considerable, but quite unknown, non-linear variation of ΔP with N was also expected for the outer part of the (3*p*) wave function. As a guess, ΔP for K^+ was taken as one-quarter of the way from Ca^{++} to Cl^-, both for (3*s*) and (3*p*). This estimate proved to be successful beyond expectation, the maximum differences between the first estimates of ΔP and the results of the final approximation being as follows:

(1*s*)	(2*s*)	(2*p*)	(3*s*)	(3*p*)
0·002	0·001	0·002	0·005	0·005

Similar interpolations for the values of $\epsilon_{nl,nl}$ were also very successful.

Two approximations were first made to the solution of the (3*s*) and (3*p*) wave functions, assuming the estimates of the others to be correct. Then solutions for (1*s*), (2*s*), (2*p*) were evaluated; these confirmed the estimates very closely, as shown by the results just quoted; then a final solution for the (3*s*), (3*p*) wave functions was evaluated.

These results showed that for (1*s*), (2*s*), (2*p*) the process of interpolation of ΔP and of $\epsilon_{nl,nl}$, linearly in N, was reliable to the accuracy required, and, for Ar, calculations were carried out for the (3*s*) and (3*p*) wave functions only. For the interpolation to provide the first estimates of these wave functions, ΔP for Ar was taken as one-third of the way from K^+ to Cl^-; two approximations were sufficient to give the final results. The greatest differences between the first estimates of ΔP and the results of the final approximation were 0·003, both for (3*s*) and (3*p*).

4. RESULTS AND DISCUSSION

The results for K^+ are given in Tables I (normalized wave functions and total $2Z_p$) and III (radial charge distribution), and corresponding results for Ar are given in Tables II and IV. Tables of contributions to Z, corresponding to the similar tables for other atoms, are not given here, since the main use of these contributions to Z has been in forming initial estimates for the solution of Fock's equations; as already explained, it seems better to start from estimates of the radial wave function P, so that these contributions to Z will not be required in future work. The function $2Z_p$ is likely to be required, however; $2Z_p$ for K^+ will be wanted in any calculations of the wave functions for the optical terms of neutral K, and $2Z_p$ for Ar will be wanted in any calculation of electron scattering.

The differences ΔP between the normalized radial wave functions calculated with and without exchange are shown in fig. 1, as already mentioned. The corresponding differences for Cu^+ are also shown, though since they refer to a configuration with the $(3d)^{10}$ group in addition to the groups of the argon-like configuration of Cl^- to Ca^{++}, they would not be expected to agree quantitatively with the results for the argon-like structures.

For $(2s)$ and $(2p)$ the use of the plot of ΔP against r/r_m brings the results for the structures Cl^- to Ca^{++} into close agreement, which is not improved by multiplication by the factor $(N - s_2)^{\frac{1}{2}}$ suggested by the argument of § 2, though the agreement of the results for Cu^+ with them is improved by the use of such a factor. For $(3s)$ and $(3p)$ the agreement is also good for $r/r_m < 1$, but not outside this value; this is not surprizing, since $(3s)^2$ and $(3p)^6$ are the outermost groups of the argon-like configuration, and for the outer part of the wave functions the approximation of regarding the mutual interactions between electrons as small compared with the interaction between an electron and the nucleus, on which the argument of § 2 is based, is a very crude one.

The results for K^+ and Ar confirm the unexpected decrease of ΔP with decreasing N for $(3s)$, for large r, already noted in the comparison of the results for Cl^- and Ca^{++}, which at first sight seemed to suggest an error in one of these results. A qualitative explanation of this decrease has already been given (D. R. and W. Hartree 1936*b*, p. 58).

One application of the present results, namely, to the evaluation of the X-ray scattering factors for K^+, has already been made by Brindley and Ridley (1938).

TABLE I. K$^+$. NORMALIZED RADIAL WAVE FUNCTIONS $P_N(nl\,|\,r)$ CALCULATED WITH EXCHANGE, AND T = TOTAL $2Z_p$

Table of $P_N(nl\,|\,r)/r^{l+1}$

r	(1*s*)	(2*s*)	(2*p*)	(3*s*)	(3*p*)
0·000	161·7	46·41	213·8	15·29$_5$	65·30
0·005	147·0	42·17	203·9	13·89$_5$	62·28
0·010	133·7	38·24	194·5$_5$	12·59$_5$	59·41
0·015	121·6$_5$	34·60	185·6$_5$	11·39	56·68
0·020	110·6$_5$	31·23	177·2	10·27	54·09

Table of P_N and T

r	(1*s*)	(2*s*)	(2*p*)	(3*s*)	(3*p*)	T
0·000	0	0	0	0	0	38·00
0·005	0·735	0·211	0·005	0·069$_5$	0·001$_5$	37·25$_5$
0·010	1·337	0·382$_5$	0·019$_5$	0·126	0·006	36·52$_5$
0·015	1·825	0·519	0·042	0·171	0·013	35·81
0·020	2·213	0·624$_5$	0·071	0·205$_5$	0·021$_5$	35·12
0·03	2·750	0·757	0·145$_5$	0·248	0·044$_5$	33·82
0·04	3·039	0·803	0·236	0·262	0·072	32·63
0·05	3·150	0·784	0·337	0·254	0·102$_5$	31·54
0·06	3·136	0·716	0·444	0·230	0·134$_5$	30·54
0·07	3·037	0·612	0·554	0·194	0·167	29·61
0·08	2·883	0·483	0·663	0·149	0·200	28·74$_5$
0·09	2·695	0·336	0·770	0·098	0·231	27·93
0·10	2·490	0·179	0·872	0·044	0·260	27·15$_5$
0·12	2·068	−0·145	1·062	−0·066	0·313	25·71
0·14	1·673	−0·458	1·225	−0·171	0·355	24·36$_5$
0·16	1·328	−0·744	1·359	−0·264	0·387	23·11$_5$
0·18	1·039	−0·992	1·464	−0·341	0·407	21·95
0·20	0·805	−1·200	1·543	−0·402	0·417	20·86$_5$
0·22	0·618	−1·366	1·596	−0·445	0·417	19·86
0·24	0·472	−1·494	1·628	−0·471	0·408	18·93
0·26	0·358	−1·586	1·640	−0·481	0·391	18·08
0·28	0·271	−1·647	1·635	−0·478	0·367	17·29
0·30	0·204	−1·680	1·616	−0·462	0·337	16·57$_5$
0·35	0·100	−1·667	1·523	−0·379	0·242	15·02$_5$
0·40	0·049	−1·562	1·388	−0·252	0·128	13·76$_5$
0·45	0·023$_5$	−1·408	1·234	−0·101	+0·005	12·71$_5$
0·50	0·011$_5$	−1·233	1·077	+0·058	−0·120	11·81$_5$
0·55	0·005$_5$	−1·058	0·926	0·216	−0·240	11·01$_5$
0·60	0·002$_5$	−0·892	0·787	0·364	−0·354	10·29
0·7	0·000$_5$	−0·612	0·552	0·616	−0·549	8·98$_5$
0·8	—	−0·405	0·377	0·798	−0·697	7·83$_5$
0·9	—	−0·262	0·253	0·912	−0·799	6·81$_5$
1·0	—	−0·167	0·167	0·969	−0·860	5·93
1·1	—	−0·105	0·109$_5$	0·981	−0·889	5·17$_5$
1·2	—	−0·066	0·071$_5$	0·961	−0·892	4·54$_5$

TABLE I (*continued*)

Table of P_N and T

r	$(1s)$	$(2s)$	$(2p)$	$(3s)$	$(3p)$	T
1·4	—	−0·026	0·030	0·860	−0·845	3·598
1·6	—	−0·010$_5$	0·013	0·725	−0·759	2·982
1·8	—	−0·004	0·005$_5$	0·589	−0·659	2·595
2·0	—	−0·001$_5$	0·002$_5$	0·466	−0·559	2·357
2·2	—	−0·005$_5$	0·001	0·362	−0·466	2·212
2·4	—	—	0·000$_5$	0·277	−0·383	2·125
2·6	—	—	—	0·210	−0·312	2·073
2·8	—	—	—	0·157	−0·252	2·042
3·0	—	—	—	0·117	−0·203	2·025
3·2	—	—	—	0·086$_5$	−0·162	2·014
3·4	—	—	—	0·064	−0·128	2·007
3·6	—	—	—	0·047	−0·101$_5$	2·004
3·8	—	—	—	0·034	−0·080	2·002
4·0	—	—	—	0·025	−0·062$_5$	2·001
4·5	—	—	—	0·011	−0·033$_5$	—
5·0	—	—	—	0·005	−0·018	—
5·5	—	—	—	0·002	−0·009$_5$	—
6·0	—	—	—	0·001	−0·005	—
6·5	—	—	—	0·000$_5$	−0·002$_5$	—
7·0	—	—	—	—	−0·001$_5$	—
7·5	—	—	—	—	−0·000$_5$	—
$\epsilon_{nl,nl}$	267·55	29·44	23·49	3·930	2·341	—

TABLE II. NEUTRAL Ar. NORMALIZED RADIAL WAVE FUNCTIONS $P_N(nl\,|\,r)$ CALCULATED WITH EXCHANGE AND T = TOTAL $2Z_p$

NOTE. Only the $(3s)$, $(3p)$ wave functions have been calculated by numerical integration of Fock's equations; the $(1s)$, $(2s)$, $(2p)$, wave functions have been calculated by interpolation of ΔP between the values for Ca^{++}, K^{+} and Cl^{-}; see text.

Table of $P_N(nl\,|\,r)/r^{l+1}$

r	$(1s)$	$(2s)$	$(2p)$	$(3s)$	$(3p)$
0	148·8	42·25	182·3$_5$	13·21	50·97
0·005	136·0	38·58	174·3$_5$	12·06$_5$	48·73
0·010	124·3	35·17	166·8	10·99$_5$	46·60
0·015	113·6$_5$	32·00	159·6	10·00	44·57
0·020	103·9$_5$	29·04	152·7	9·07$_5$	42·63

Table of P_N and T

r	$(1s)$	$(2s)$	$(2p)$	$(3s)$	$(3p)$	T
0	0	0	0	0	0	36·00
0·005	0·680	0·193	0·004$_5$	0·060$_5$	0·001	35·30$_5$
0·010	1·243	0·352	0·016$_5$	0·110	0·004$_5$	34·62
0·015	1·706	0·480	0·036	0·150	0·010	33·95
0·020	2·079	0·581	0·061	0·181$_5$	0·017	33·30$_5$

TABLE II (*continued*)

Table of P_N and T

r	(1*s*)	(2*s*)	(2*p*)	(3*s*)	(3*p*)	T
0·03	2·608	0·713	0·126	0·222	0·035	32·08
0·04	2·910	0·769	0·205	0·239	0·057	$30{\cdot}95_5$
0·05	3·047	0·764	0·294	0·236	0·082	29·92
0·06	3·063	0·715	0·389	0·219	0·108	28·97
0·07	2·996	0·631	0·487	0·191	0·135	$28{\cdot}08_5$
0·08	2·873	0·522	0·586	0·155	0·162	27·26
0·09	2·712	0·395	0·684	0·113	0·188	$26{\cdot}48_5$
0·10	2·528	0·258	0·780	0·068	0·214	25·75
0·12	2·141	−0·035	0·957	−0·027	0·260	24·37
0·14	1·767	−0·326	1·114	−0·120	0·298	23·10
0·16	1·433	−0·597	1·247	−0·205	0·329	21·91
0·18	1·147	−0·840	1·355	−0·278	0·351	$20{\cdot}79_5$
0·20	0·905	−1·048	1·440	−0·338	0·365	19·75
0·22	0·708	−1·221	1·503	−0·384	0·371	$18{\cdot}77_5$
0·24	0·550	−1·361	1·546	−0·416	0·369	17·87
0·26	0·425	−1·468	1·571	−0·436	0·361	17·03
0·28	0·327	−1·546	1·580	−0·443	0·347	$16{\cdot}25_5$
0·30	0·251	−1·597	1·575	−0·439	0·327	$15{\cdot}53_5$
0·35	0·128	−1·631	1·516	−0·389	0·261	$13{\cdot}97_5$
0·40	0·065	−1·570	1·411	−0·296	0·175	$12{\cdot}70_5$
0·45	0·032	−1·448	1·282	−0·177	0·078	$11{\cdot}64_5$
0·50	0·016	−1·298	1·143	−0·044	−0·023	10·75
0·55	$0{\cdot}007_5$	−1·140	1·004	+0·093	−0·124	$9{\cdot}96_5$
0·60	$0{\cdot}003_5$	−0·983	0·870	0·226	−0·222	9·26
0·7	$0{\cdot}000_5$	−0·703	0·637	0·465	−0·399	$8{\cdot}01_5$
0·8	—	−0·486	0·453	0·654	−0·545	$6{\cdot}90_5$
0·9	—	−0·327	0·316	0·789	−0·656	5·91
1·0	—	−0·216	0·218	0·873	−0·735	5·02
1·1	—	−0·142	0·148	0·915	−0·786	4·23
1·2	—	−0·092	0·100	0·924	−0·813	3·54
1·4	—	−0·038	0·045	0·875	−0·814	2·442
1·6	—	−0·016	0·020	0·779	−0·771	1·659
1·8	—	$-0{\cdot}006_5$	0·009	0·666	−0·705	1·116
2·0	—	$-0{\cdot}002_5$	0·004	0·555	−0·629	0·745
2·2	—	−0·001	0·002	0·454	−0·553	0·495
2·4	—	$-0{\cdot}000_5$	0·001	0·366	−0·481	0·328
2·6	—	—	$0{\cdot}000_5$	0·293	−0·414	0·217
2·8	—	—	—	0·232	−0·355	0·144
3·0	—	—	—	0·183	−0·302	0·095
3·2	—	—	—	0·143	−0·256	0·063
3·4	—	—	—	0·112	−0·216	0·041
3·6	—	—	—	0·087	−0·182	0·027
3·8	—	—	—	0·068	−0·153	0·017
4·0	—	—	—	0·052	−0·128	0·011

TABLE II (*continued*)

Table of P_N and T

r	$(1s)$	$(2s)$	$(2p)$	$(3s)$	$(3p)$	T
4·5	—	—	—	0·027	−0·082	0·004
5·0	—	—	—	0·014	−0·051$_5$	0·001
5·5	—	—	—	0·007$_5$	−0·032	—
6·0	—	—	—	0·004	−0·020	—
6·5	—	—	—	0·002	−0·012$_5$	—
7·0	—	—	—	0·001	−0·007$_5$	—
7·5	—	—	—	0·000$_5$	−0·004$_5$	—
8·0	—	—	—	—	−0·003	—
9·0	—	—	—	—	−0·001	—
10·0	—	—	—	—	−0·000$_5$	—
$\epsilon_{nl\,nl}$	237·2	24·6$_5$	19·1$_5$	2·555	1·181	—

TABLE III. K$^+$. CHARGE DISTRIBUTION

Total radial charge density in atomic units $U(r) = \Sigma_{nl}\, 2(2l+1)\, P_N^2(nl\,|\,r)$ calculated by self-consistent field (*a*) without exchange, (*b*) with exchange, and the difference $\delta U(r)$.

r	$U(r)$ (a)	$U(r)$ (b)	$\delta U(r)$ $(b)-(a)$	r	$U(r)$ (a)	$U(r)$ (b)	$\delta U(r)$ $(b)-(a)$
0	0	0	0	0·50	10·38$_5$	10·10	−0·28$_5$
0·005	1·18	1·18	0	0·55	7·99	7·82	−0·17
0·010	3·90	3·90	0	0·60	6·31	6·32	+0·01
0·015	7·26	7·26$_5$	0·00$_5$	0·7	4·75$_5$	5·14$_5$	0·39
0·020	10·68$_5$	10·70$_5$	0·02	0·8	4·72$_5$	5·36$_5$	0·64
0·03	16·51$_5$	16·53	0·01$_5$	0·9	5·26$_5$	6·01	0·74$_5$
0·04	20·25	20·25$_5$	0·00$_5$	1·0	5·83	6·54$_5$	0·71$_5$
0·05	21·94$_5$	21·95	0·00$_5$	1·1	6·16$_5$	6·76$_5$	0·60
0·06	22·07	22·09	0·02	1·2	6·21	6·65$_5$	0·44$_5$
0·07	21·24	21·28	0·04	1·4	5·63$_5$	5·76$_5$	+0·13
0·08	19·93	20·00$_5$	0·07$_5$	1·6	4·62	4·51	−0·11
0·09	18·52	18·64$_5$	0·12$_5$	1·8	3·55$_5$	3·30	−0·25$_5$
0·10	17·26$_5$	17·44	0·17$_5$	2·0	2·63	2·31	−0·32
0·12	15·65	15·96	0·31	2·2	1·89	1·56$_5$	−0·32$_5$
0·14	15·38$_5$	15·82$_5$	0·44	2·4	1·33$_5$	1·03$_5$	−0·30
0·16	16·19	16·74$_5$	0·55$_5$	2·6	0·927	0·674	−0·253
0·18	17·57$_5$	18·22	0·64$_5$	2·8	0·637	0·432	−0·205
0·20	19·14$_5$	19·83	0·68$_5$	3·0	0·435	0·274	−0·161
0·22	20·54	21·23	0·69	3·2	0·292	0·172	−0·122
0·24	21·60$_5$	22·25	0·64$_5$	3·4	0·196	0·107	−0·089
0·26	22·22	22·79$_5$	0·57$_5$	3·6	0·130	0·066	−0·064
0·28	22·39$_5$	22·87	0·47$_5$	3·8	0·088	0·040	−0·048
0·30	22·15	22·50	0·35	4·0	0·057	0·025	−0·032
0·35	20·08$_5$	20·12$_5$	+0·04	4·5	0·020	0·007	−0·013
0·40	16·85$_5$	16·67	−0·18$_5$	5·0	0·007	0·002	−0·005
0·45	13·42$_5$	13·12$_5$	−0·30	5·5	0·002	—	−0·002

TABLE IV. NEUTRAL Ar. CHARGE DISTRIBUTION

Total radiation charge density in atomic units, $U(r) = \Sigma_{nl}\, 2(2l+1)\, P^2_N(nl|r)$ calculated by self-consistent field (*a*) without exchange, (*b*) with exchange and the difference $\delta U(r)$.

r	$U(r)$ (*a*)	$U(r)$ (*b*)	$\delta U(r)$ (*b*) − (*a*)
0	0	0	0
0·005	1·00$_5$	1·00$_5$	0·00
0·010	3·36	3·36	0·00
0·015	6·32$_5$	6·33	0·00$_5$
0·020	9·40$_5$	9·41$_5$	0·01
0·03	14·80$_5$	14·82	0·01$_5$
0·04	18·49$_5$	18·50$_5$	0·01
0·05	20·39$_5$	20·40$_5$	0·01
0·06	20·84$_5$	20·86	0·01$_5$
0·07	20·33	20·36	0·03
0·08	19·26	19·32	0·06
0·09	17·97	18·07	0·10
0·10	16·70	16·84$_5$	0·14$_5$
0·12	14·82	15·07	0·25
0·14	14·09	14·46$_5$	0·37$_5$
0·16	14·39	14·88	0·49
0·18	15·37$_5$	15·95$_5$	0·58
0·20	16·67	17·30	0·63
0·22	17·99	18·66	0·67
0·24	19·15	19·81$_5$	0·66$_5$
0·26	20·01	20·64$_5$	0·63$_5$
0·28	20·50$_5$	21·08	0·57$_5$
0·30	20·64$_5$	21·13$_5$	0·49
0·35	19·63	19·85	0·22
0·40	17·27$_5$	17·24	−0·03$_5$
0·45	14·38$_5$	14·16	−0·22$_5$
0·50	11·51	11·21	−0·30
0·55	9·00$_5$	8·74$_5$	−0·26
0·60	7·04	6·88	−0·16
0·7	4·67	4·81$_5$	+0·14$_5$
0·8	3·90$_5$	4·34$_5$	0·44
0·9	4·01$_5$	4·64$_5$	0·63
1·0	4·45	5·14$_5$	0·69$_5$
1·1	4·86$_5$	5·54	0·67$_5$
1·2	5·13	5·74$_5$	0·61$_5$
1·4	5·13	5·52	0·39
1·6	4·62	4·78	0·16
1·8	3·89	3·86$_5$	−0·02$_5$
2·0	3·14	2·99	−0·15
2·2	2·48	2·24$_5$	−0·23$_5$
2·4	1·929	1·653	−0·276
2·6	1·483	1·199	−0·284
2·8	1·133	0·862	−0·271
3·0	0·861	0·613	−0·248
3·2	0·651	0·434	−0·217
3·4	0·492	0·306	−0·186
3·6	0·371	0·215	−0·156
3·8	0·278	0·150	−0·128
4·0	0·209	0·104	−0·105
4·5	0·101	0·041	−0·060
5·0	0·049	0·016	−0·033
5·5	0·023	0·006	−0·017
6·0	0·011	0·002	−0·009
6·5	0·005	0·001	−0·004
7·0	0·002	—	−0·002
7·5	0·001	—	−0·001

5. DIAMAGNETIC SUSCEPTIBILITIES

As mentioned in the previous paper on Ca (D. R. and W. Hartree 1938), one of the reasons for undertaking the solution of Fock's equation for K^+ and Ar was to have results for the calculation of diamagnetic susceptibilities, in the hope that a comparison of the difference between the calculated and observed susceptibilities for Ar, and for the difference $K^+ - Na^+$, might throw some light on the origin of these differences for ions. Some discussion of these differences has already been given by Hoare and Brindley (1937).

The calculated susceptibilities for Ar and K$^+$, and contributions to them for the different groups, are shown in Table V, and the totals are compared with the observed values. For argon, the "observed" values of $-\chi \times 10^6$ recorded by different investigators range from 18·1 to 25·3; the value here quoted in Table V is that of Mann (1936).* Earlier values have been discussed fully by Stoner (1934, p. 262); of them, Havens' (1933) value 19·2 is regarded by Stoner as the most reliable, though perhaps 2 or 3 % low, as the measurements were made relative to oxygen, for which a possibly low susceptibility was taken when converting the measurements to absolute values of susceptibility; with such a correction, Havens' value agrees closely with Mann's. For K$^+$ the observed values is that given by Brindley and Hoare (1935).

TABLE V. AR AND K$^+$. DIAMAGNETIC SUSCEPTIBILITIES

	Ar		K$^+$	
Group	$\overline{r^2}$	$2(2l+1)\,\overline{r^2}$	$\overline{r^2}$	$2(2l+1)\,\overline{r^2}$
With exchange: $(1s)^2$	0·010	0·02	0·009	0·02
$(2s)^2$	0·201	0·40	0·117	$0{\cdot}23_5$
$(2p)^6$	0·174	$1{\cdot}04_5$	0·151	$0{\cdot}90_5$
$(3s)^2$	2·348	$4{\cdot}69_5$	1·881	$3{\cdot}76_5$
$(3p)^6$	3·312	19·87	2·414	$14{\cdot}48_5$
S.c.f. with exchange	Total	26·03		19·41
	$-\int_0^\infty r^2 \delta U(r)\, dr$	5·15		2·65
S.c.f. without exchange	Total	31·18		22·06
Calc. with exchange	$-\chi \times 10^6$	20·6		15·3
Calc. without exchange	$-\chi \times 10^6$	24·6		17·4
Observed	$-\chi \times 10^6$	19·5		14·6

It will be seen that for argon the use of wave functions calculated with exchange leads to a marked improvement in the agreement between the calculated and observed susceptibilities; the calculated value still appears somewhat too high, but the difference is perhaps not much greater than the possible error of the observed value quoted.

"Observed" values of susceptibilities for single ions are affected by the uncertainty arising from the step of dividing the susceptibility of a salt (either crystalline or in solution) into contributions from its positive and negative ions; a constant added to the susceptibilities of all positive ions and subtracted from those of all negative ions would have no effect on those of salts. On the other hand, the difference between the susceptibilities of two positive ions is unaffected by this uncertainty, and it is more profitable therefore to compare the calculated and observed values of such a difference than the values of the susceptibilities themselves. The present results for K^+, the value for Na^+ (D. R. and W. Hartree 1936*b*, p. 61) calculated from Fock and Petrashen's wave functions make such a comparison possible; the figures are

---

* We are indebted to Professor Hassé and to Dr Stoner for this reference.

$$\begin{aligned} -(\chi_{K^+}-\chi_{Na^+})\times 10^6 &= 15{\cdot}3-4{\cdot}9 = 10{\cdot}4 \text{ calculated} \\ &= 14{\cdot}6-6{\cdot}1 = 8{\cdot}5 \text{ observed} \end{aligned}$$

(the observed values from the individual ions are taken from Brindley and Hoare (1935, Table VI)).

It will be seen that the calculated difference is appreciably greater than the observed, though the disagreement is much smaller than when the susceptibility is evaluated from wave functions calculated without exchange. Now for argon the calculated susceptibility is only very little different from the observed value, and it would be expected that the error in the approximation from which Fock's equations are derived would be relatively smaller for a positive ion than for the corresponding neutral atom; so the difference between the calculated and true values of $-(\chi_{K^+}-\chi_{Na^+})\times 10^6$ for free ions would not be expected to be greater than 1 at most, whereas the difference between $-(\chi_{K^+}-\chi_{Na^+})\times 10^6$ calculated for free ions and observed for crystals is nearly 2. Now it is known, both from diamagnetic susceptibilities and from packing considerations, that the outer part of the charge distribution of the Cl^- ion must be appreciably distorted in a crystal as compared to the distribution in a free ion. The present comparison of values of $-(\chi_{K^+}-\chi_{Na^+})$ suggests that this distortion is also appreciable for the smaller and more tightly bound distribution of a positive ion. The assumption of appreciable distortion of the ions on packing into a crystal lattice does not, however, seem to explain the rather surprising result that for Na^+ the calculated $-\chi$ is smaller than the observed value.

6. SUMMARY

The differences between the atomic wave functions calculated by the method of the self-consistent field with and without exchange can in certain cases be plotted in such a way that the results for different atoms fall on approximately the same curve, which then can be used for interpolating between different atoms.

Wave functions so estimated, using the results already calculated for Cl^- and Ca^{++}, have been used as the basis of calculations of the self-consistent field for K^+ and Ar. For K^+ all the (nl) wave functions have been calculated; the results showed that for Ar all but the outer [$(3s)$ and $(3p)$] wave functions could be interpolated to adequate accuracy by the method mentioned.

Results are given, and values of the diamagnetic susceptibilities calculated and compared with observed values.

REFERENCES

Brindley, G. W. and Hoare, F. E. 1935 *Proc. Roy. Soc.* A, **152**, 342–53.
Brindley, G. W. and Ridley, P. 1938 *Proc. Phys. Soc.* **50**, 96–107.
Fock, V. 1930 *Z. Phys.* **61**, 126–48.
Hartree, D. R. 1934 *Proc. Roy. Soc.* A, **143**, 506–17.
Hartree, D. R. and Hartree, W. 1935 *Proc. Roy. Soc.* A, **150**, 9–33.
— — 1936*a* *Proc. Roy. Soc.* A, **154**, 588–607.
— — 1936*b* *Proc. Roy. Soc.* A, **156**, 45–62.
— — 1936*c* *Proc. Roy. Soc.* A, **157**, 490–502.
— — 1938 *Proc. Roy. Soc.* A, **164**, 167–91.
Havens, G. G. 1933 *Phys. Rev.* **43**, 992–1000.
Hoare, F. E. and Brindley, G. W. 1937 *Proc. Roy. Soc.* A, **159**, 395–409.
Mann, K. E. 1936 *Z. Phys.* **98**, 546.
Stoner, E. C. 1934 "Magnetism and Matter." London.

The problem of n bodies in general relativity theory

By Sir Arthur Eddington, F.R.S., and G. L. Clark

(*Received* 22 *March* 1938)

1. In a recent investigation of the problem of two bodies in general relativity theory, Prof. Levi-Civita (1937*b*) has reached the conclusion that the centre of gravity has a secular acceleration in the direction of the major axis of the orbit towards the periastron of the larger mass. He gives an example of a binary star in which this acceleration may become detectable in much less than a century, perhaps even in a few years.

Since the periastron slowly revolves, Levi-Civita's result implies that the binary star as a whole spontaneously describes a circle of very large radius. There seems to be no obvious ground for regarding this as impossible; it is conceivable that an unsymmetrical system may radiate momentum by gravitational waves and so experience a recoil. But the behaviour is so peculiar as to inspire doubt; and we determined to re-examine the problem with a view either to detecting an error or to obtaining further light on the nature of the phenomenon.

The conclusion of our investigation is that there is *no* secular acceleration of the centre of gravity (§ 8).

Levi-Civita's algebraic calculations are not given in sufficient detail for us to locate the source of the divergence. It may be pointed out, however, that the null result is due to the mutual cancelling of a number of terms, so that a numerical slip in the coefficient of one of the terms may be responsible. Our first calculation gave a positive result, with a coefficient differing from Levi-Civita's but of the same order of magnitude. This was based on de Sitter's general formula for the line-element due to n bodies. We discovered, however, that de Sitter had made an error of theory in the calculation of one of his terms (§ 2); and when this was corrected, the secular acceleration became zero. De Sitter's investigation is quoted by Levi-Civita in his paper; but we have learnt by correspondence with him that he did not take over the formula containing the error.

For a system of two bodies our result confirms the usual assumption that the system as a whole has no spontaneous acceleration. We have not been able to generalize the proof to apply to n bodies. The obstacle is that, whereas in the two-body problem we use the approximate Newtonian solution for evaluating small terms, in the n-body problem the corresponding Newtonian solution does not exist.

Since de Sitter's formula for the line-element has frequently been referred to in the past twenty years, its amendment is of importance apart from the particular application to Levi-Civita's problem. In particular the calculation of the equivalent mass of a system of particles now comes into line with the theory of the equivalent mass of a system of continuous matter developed by Tolman, Whittaker and others (§ 4). Previously we had been puzzled by a failure to reconcile them.

2. The first discussion of the field of n bodies was given by Droste (1916). Subsequently de Sitter (1916), making a minor alteration, used the result to investigate problems in gravitational astronomy. The error, mentioned in § 1, occurs on pp. 156–7 of de Sitter's paper. He has obtained (equation (65))

$$\nabla^2\gamma = \kappa\rho, \tag{2·1}$$

and writes the solution as

$$\gamma = -\frac{\kappa}{4\pi}\int\frac{\rho\, dV}{r}. \tag{2·2}$$

The equations (2·1) and (2·2) are most familiar to us as determining a potential in Euclidean space; but the solution is, of course, purely analytical, and does not depend on a geometrical representation of the variables x_1, x_2, x_3 contained in ∇^2. It must, however, be remembered that when, as here, the space is non-Euclidean, r in (2·2) is a fictitious distance calculated as though x_1, x_2, x_3 were Euclidean rectangular co-ordinates, and dV is similarly the fictitious volume $dx_1 dx_2 dx_3$ of the element. De Sitter seems to have overlooked the latter point, and to have taken dV to be the natural volume; thus he assumes that "in a system of co-ordinates in which a body would be at rest $\int\rho\, dV$ would be the mass" (de Sitter's ρ is the invariant density T). The natural volume is $\sqrt{(g_{11}g_{22}g_{33})}\,.\,dx_1 dx_2 dx_3$, so that the correct formula for the mass is

$$(g_{11}g_{22}g_{33})^{\frac{1}{2}}\int\rho\, dV, \tag{2·3}$$

subject to the usual correction (duly inserted by de Sitter) for change of mass with velocity.

3. It is well known that even the problem of two bodies has not been solved exactly in general relativity theory. The order of approximation aimed at by de Sitter and by Levi-Civita is to determine the accelerations

correctly to the squares of the potentials or equivalently to the fourth powers of the velocities. Treating the Newtonian potential as of the first order, the order of the terms required to be retained in the $g_{\mu\nu}$ is found to be as follows:

$$g_{44},\ \text{order } 2; \quad g_{14}, g_{24}, g_{34},\ \text{order } \tfrac{3}{2}; \quad \text{the others, order } 1.$$

The usual method of first approximation (Eddington 1924, § 57) gives ($\mu = 1, 2, 3$)

$$\left.\begin{aligned} g_{\mu\mu} = -1+\gamma_0, \quad g_{44} = 1+\gamma_0, \quad g_{12}, g_{23}, g_{31} = 0, \\ g_{\mu 4} = 4\,\Sigma_i\, m_i (v_\mu)_i / r_i, \end{aligned}\right\} \tag{3·1}$$

where
$$\gamma_0 = -2\,\Sigma_i\, m_i/r_i. \tag{3·2}$$

The units are such that the constant of gravitation and the velocity of light are unity; $(v_\mu)_i$ is the velocity of the ith body, and r_i is its distance (calculated as for Euclidean co-ordinates) from the point considered.

This approximation is of the required accuracy except for g_{44}. To obtain the higher approximation $g_{44} = 1+\gamma$, we use the exact equation furnished by

$$G_{44} = -8\pi(T_{44} - \tfrac{1}{2}g_{44}T),$$

and simplify it by inserting the first approximation for the $g_{\mu\nu}$ other than g_{44}, and also for g_{44} when operated on by $\partial/\partial t$. ($\partial g_{44}/\partial t$ consists of terms of the form $\dfrac{\partial g_{44}}{\partial (x_\mu)_i}\cdot\dfrac{(\partial x_\mu)_i}{\partial t}$, and the second factor is of order $\tfrac{1}{2}$.) The equation for γ is found to be (de Sitter 1916, equation (69))

$$\Sigma' \frac{\partial^2 \gamma}{\partial x_\mu^2} = 8\pi(\rho_0 + 2\rho_0 v^2) + \frac{\partial^2 \gamma_0}{\partial t^2} + \Sigma' \left(\frac{\partial \gamma_0}{\partial x_\mu}\right)^2, \tag{3·3}$$

where Σ' denotes summation for x_1, x_2, x_3; ρ_0 is the invariant density (T), and v the velocity of matter at the point considered. The solution (obtained by de Sitter) is

$$\gamma = \tfrac{1}{2}\gamma_0^2 + \Sigma_i\left(-\frac{2q_i}{r_i} - \frac{4m_i v_i^2}{r_i} - \frac{4m_i}{r_i}\Sigma_j \frac{m_j}{\Delta_{ij}} - m_i \frac{\partial^2 r_i}{\partial t^2}\right), \tag{3·4}$$

where Δ_{ij} is the distance between the ith and jth bodies (the value $j = i$ being omitted in the j-summation), and

$$q_i = \int \rho_0\, dV$$

taken over the ith body. De Sitter, taking dV to be the natural volume, gives

$$q_i = m_i(1 - v_i^2)^{\frac{1}{2}}.$$

But taking account of the factor $(g_{11}g_{22}g_{33})^{\frac{1}{2}}$ in (2·3), the correct result is

$$q_i = m_i(1 + \tfrac{3}{2}(\gamma_0)_i - \tfrac{1}{2}v_i^2) \tag{3·5}$$

to the required approximation. The value of $(\gamma_0)_i$, i.e. γ_0 evaluated at the ith body, is

$$(\gamma_0)_i = -2\,\Sigma_j m_j/\Delta_{ij} \quad (j \neq i).$$

Hence by (3·4) and (3·5)

$$\gamma = \gamma_0 + \tfrac{1}{2}\gamma_0^2 + \Sigma_i\left(-\frac{3m_i v_i^2}{r_i} + \frac{2m_i}{r_i}\Sigma_j \frac{m_j}{\Delta_{ij}} - m_i\frac{\partial^2 r_i}{\partial t^2}\right). \tag{3·6}$$

This differs from de Sitter's value (p. 159, equation (70)) by having the coefficient $+2$ instead of -4 in the middle term in the bracket.

4. If we adopt a co-ordinate system in which the centre of mass of the system of n particles is at rest at the instant considered, so that $\Sigma_i m_i(v_\mu)_i = 0$, the $g_{\mu\nu}$ found in § 3 give for the line-element at great distances

$$ds^2 = (-1 + \gamma_0)(dx_1^2 + dx_2^2 + dx_3^2) + (1 + \gamma)\,dx_4^2. \tag{4·1}$$

We define the mass of a system (at the instant considered) to be the mass M of an equivalent particle which gives the same line-element at great distances.* The line-element of the particle in isotropic co-ordinates is (Eddington 1924, § 43)

$$ds^2 = -(1 + M/2r)^4(dx_1^2 + dx_2^2 + dx_3^2) + \frac{(1 - M/2r)^2}{(1 + M/2r)^2}dx_4^2.$$

Writing $\gamma_M = -2M/r$, and expanding in powers of γ_M, this gives

$$ds^2 = (-1 + \gamma_M + \ldots)(dx_1^2 + dx_2^2 + dx_3^2) + (1 + \gamma_M + \tfrac{1}{2}\gamma_M^2 + \ldots)\,dx_4^2. \tag{4·2}$$

Comparing (4·1) and (4·2) we see that γ must approach $\gamma_M + \frac{1}{2}\gamma_M^2$ at great distances.

We have retained the second order term $\frac{1}{2}\gamma_M^2$ in g_{44} because it is the counterpart of the term $\frac{1}{2}\gamma_0^2$ in the expression (3·6) for γ, and shows that the equivalence of the fields of the system and of the particle extends to the second order. But for the determination of M, the first order terms suffice; γ_0^2 is then negligible, and γ must approach $-2M/r$ when the r_i become large and equal to r. We then obtain by (3·6)

$$M = \Sigma_i m_i + \tfrac{3}{2}\Sigma_i m_i v_i^2 - \Sigma_i\Sigma_j\frac{m_i m_j}{\Delta_{ij}} + \tfrac{1}{2}r\,\Sigma_i m_i\frac{\partial^2 r_i}{\partial t^2}. \tag{4·3}$$

---

* The definition will be made more precise later. Meanwhile we contemplate a distance large from the point of view of our approximations, but not so large as to introduce an important time-lag of the potentials.

In the last term, $\Sigma_i m_i \partial^2 r_i/\partial t^2$ vanishes if the centre of mass of the system has zero acceleration; but we must not assume that this condition is fulfilled. Even if there is no secular acceleration, the centre of mass has usually an instantaneous acceleration. It is understood that the centre of mass is defined in the elementary way, using the rest-masses of the particles; a more complicated definition would be required if we wished to define a point which remains fixed during the motion, bearing in mind the relativistic variations of the masses of the particles.

Accordingly the term $r\Sigma m_i \partial^2 r_i/\partial t^2$ does not in general vanish. It can be interpreted as the correction for retardation of the potentials contained in γ_0. Denoting retarded values by square brackets, the expansion of a retarded potential in terms of present values is (Eddington 1924, p. 253, equation (4))

$$\frac{m}{[r(1-v_r)]} = \frac{m}{r} + \tfrac{1}{2} m \frac{d^2 r}{dt^2} - \ldots. \tag{4·4}$$

Thus in (3·6), $\gamma_0 - \Sigma m_i \partial^2 r_i/\partial t^2$ could be written as $[\gamma_0]$.

We can therefore remove the last term of (4·3) by exercising a little more care in defining the equivalent mass of a system. First, we recognize that the comparison of fields at great distances determines the mass of the system at an appropriately antedated instant. Secondly, it is to be understood that the equivalent particle has the same acceleration, as well as the same velocity, as the centre of mass of the system. Then (4·2) must be corrected for the retardation of potential of the accelerated particle, or equivalently the retardation terms must be removed from (4·1). The mass is then

$$M = \Sigma_i m_i + \tfrac{3}{2}\Sigma_i m_i v_i^2 - \Sigma_i \Sigma_j \frac{m_i m_j}{\Delta_{ij}}, \tag{4·5}$$

and the formal difficulty, caused by the divergence of (4·3) as $r \to \infty$, is avoided.

Let T, V be the kinetic and potential energies of the system. We have

$$T = \tfrac{1}{2}\Sigma_i m_i v_i^2, \quad V = -\tfrac{1}{2}\Sigma_i \Sigma_j \frac{m_i m_j}{\Delta_{ij}}.$$

Hence (4·5) becomes

$$M = \Sigma m_i + 3T + 2V. \tag{4·6}$$

It has been shown that in a system of gravitating particles (Eddington 1916, equation (4))

$$2T + V = \frac{1}{2}\frac{d^2 C}{dt^2}, \tag{4·7}$$

where C is the moment of inertia of the system about its centre of mass. The whole energy of the bodies, including rest-energy, is

$$E = \Sigma m_i + T + V. \tag{4·8}$$

Hence by (4·6), (4·7) and (4·8)

$$M = E + \frac{1}{2}\frac{d^2C}{dt^2}. \tag{4·9}$$

Thus, in general the mass of a system is not equal to the energy of the bodies composing it unless its moment of inertia is steady.

The relation $M = E$ appears in the work of Tolman, Whittaker and others on systems composed of continuous matter; but the supplementary term $\frac{1}{2}d^2C/dt^2$ in non-static systems appears to be new.

5. The $g_{\mu\nu}$ of the field having been determined, the track of a body in the field is given by the geodesic

$$\frac{d^2x_\mu}{ds^2} + \{\alpha\beta, \mu\}\frac{dx_\alpha}{ds}\frac{dx_\beta}{ds} = 0. \tag{5·1}$$

Before applying this to one of the bodies of the system we must remove the self-contribution of that body to the field. As Levi-Civita has pointed out, very delicate questions arise as to this "effacing principle".

In the first place, the field is non-linear and is therefore not a simple sum of contributions from the different bodies, so that the rule that the contribution of one body is to be removed is ambiguous. The adopted procedure is determined by the following considerations. When, after performing the differentiations in (5·1), we identify the current co-ordinates x_μ with the co-ordinates $(x_\mu)_i$ of one of the bodies of the system, r_i vanishes and a number of terms becomes infinite. It is clearly these terms which must be effaced. Therefore in (3·6) we omit the body itself in the i summation; but it is to be retained in the j summation in the double summation term.

The justification for this different treatment of the i and j summations is that the i-infinities are spurious, being consequences of treating a finite body as a particle. They should in any case be replaced by volume integrals. In the j summation, on the other hand, the treatment of the body as a particle remains valid. The effacement is therefore primarily a correction of an invalid approximation; but the assumption that the terms actually vanish requires justification.

This argument removes any possible ambiguity as to how the effacement is to be carried out. On the broader question as to how far the effacement principle can be logically or physically justified, we cannot enter here. It is connected with the problem of internal equilibrium of material bodies; and it is doubtful whether a final answer can be reached without a study of the combination of quantum and relativity theory. A large part of Levi-Civita's investigation (1937*a*) is concerned with it; but we do not think that the difference of our results for the secular acceleration of the centre of mass of a double star arises at this point, for he concludes that, when there are two bodies only, it is certainly justified.

6. The equation resulting from (5·1) is*

$$\begin{aligned}\frac{d^2x_\mu}{dt^2}+\Sigma_i\frac{m_i(x_\mu-x_{\mu i})}{r_i^3} &= 4\,\Sigma_i\,\Sigma_j\,m_i m_j\frac{x_\mu-x_{\mu i}}{r_i^3 r_j}+4\,\Sigma'\,\Sigma_i\frac{m_i(x_\alpha-x_{\alpha i})}{r_i^3}v_\mu v_\alpha \\ &\quad -\Sigma_i\frac{m_i(x_\mu-x_{\mu i})}{r_i^3}\left(v^2+2v_i^2-\tfrac{3}{2}\epsilon_i^2+\tfrac{1}{2}r_i\eta_i\right) \\ &\quad +\Sigma_i\,m_i\left\{\frac{7}{2}\frac{1}{r_i}\frac{d^2x_{\mu i}}{dt^2}+\frac{3(v_\mu-v_{\mu i})}{r_i^2}\epsilon_i+4\,\Sigma'\frac{\{(x_\mu-x_{\mu i})\,v_{\alpha i}-(x_\alpha-x_{\alpha i})\,v_{\mu i}\}\,v_\alpha}{r_i^3}\right\} \\ &\quad +\Sigma_i\frac{m_i(x_\mu-x_{\mu i})}{r_i^3}\Sigma_{(j\neq i)}\frac{m_j}{\Delta_{ij}}+m\Sigma_i\,m_i\frac{x_\mu-x_{\mu i}}{r_i^4}, \qquad (6{\cdot}1)\end{aligned}$$

where $$r_i\eta_i=\Sigma'\,(x_{\alpha i}-x_\alpha)\,d^2x_{\alpha i}/dt^2,\quad r_i\epsilon_i=\Sigma'\,(x_{\alpha i}-x_\alpha)\,v_{\alpha i}. \qquad (6{\cdot}2)$$

The summations Σ' are over the values $\alpha = 1, 2, 3$ of the co-ordinate suffix; the summations Σ are over the $n-1$ bodies of the system excluding the body whose track is being determined. The mass, co-ordinates and velocity of the latter body are denoted by m, x_μ, v_μ; its contribution has been separated from the summations and is shown explicitly in the last term. Quantities referring to the other particles carry suffixes i or j.

The above equation corresponds to (82) in de Sitter's paper. The last two terms are different on account of the emendation made in § 2. Independently of this correction, there is an error (apparently a misprint) in the coefficient of the last term in de Sitter's result.

7. We now consider a system of two bodies of masses m_1, m_2. The co-ordinate system is chosen so that the relative motion, on the Newtonian

---

* In the double summations the value $j = i$ is not excluded, except in the term where the exclusion is indicated. A simple rule (applicable throughout this paper) is to adopt the convention $\Delta_{ii} = \infty$, and retain $j = i$ in all double summations.

level of approximation is in the plane (x_1, x_2). From (6·1) we obtain for the motion of the body of mass m_2

$$\frac{d^2x_{\mu 2}}{dt^2}+\frac{m_1(x_{\mu 2}-x_{\mu 1})}{r^3}=\frac{4m_1^2(x_{\mu 2}-x_{\mu 1})}{r^4}+4m_1v_{\mu 2}\frac{\Sigma' v_{\alpha 2}(x_{\alpha 2}-x_{\alpha 1})}{r^3}$$

$$-\frac{m_1(x_{\mu 2}-x_{\mu 1})}{r^3}(v_2^2+2v_1^2-\tfrac{3}{2}\epsilon_1^2+\tfrac{1}{2}r\eta_1)$$

$$+m_1\left\{\frac{7}{2r}\frac{d^2x_{\mu 1}}{dt^2}+\frac{3(v_{\mu 2}-v_{\mu 1})\,\epsilon_1}{r^2}+\frac{4\,\Sigma'\,\{(x_{\mu 2}-x_{\mu 1})\,v_{\alpha 1}-(x_{\alpha 2}-x_{\alpha 1})\,v_{\mu 1}\}\,v_{\alpha 2}}{r^3}\right\}$$

$$+\frac{m_1m_2(x_{\mu 2}-x_{\mu 1})}{r^4}, \qquad (7{\cdot}1)$$

where r is the distance between the two bodies.

On the right-hand side of (7·1) we may substitute the approximate Newtonian values

$$\frac{d^2x_{\mu 1}}{dt^2}=\frac{m_2x_\mu}{r^3},\quad \frac{d^2x_{\mu 2}}{dt^2}=-\frac{m_1x_\mu}{r^3},\quad v_{\mu 1}=-\frac{m_2}{m}v_\mu,\quad v_{\mu 2}=\frac{m_1}{m}v_\mu,$$

$$\epsilon_1=\frac{m_2}{m}\frac{dr}{dt},\quad \epsilon_2=\frac{m_1}{m}\frac{dr}{dt},\quad r\eta_1=-\frac{m_2}{r},\quad r\eta_2=-\frac{m_1}{r},$$

$$v^2=2E+\frac{2m}{r},\quad \left(\frac{dr}{dt}\right)^2=v^2-\frac{h^2}{r^2},\quad h^2=ma(1-e^2),$$

where $m=m_1+m_2$, E is the energy of the relative motion, and v_μ and x_μ also refer to the relative motion, i.e.

$$v_\mu=v_{\mu 2}-v_{\mu 1},\quad x_\mu=x_{\mu 2}-x_{\mu 1}.$$

Equation (7·1) then reduces to

$$\frac{d^2x_{\mu 2}}{dt^2}+\frac{m_1x_\mu}{r^3}$$

$$=\frac{x_\mu}{r^3}\left\{\left(4m_1^2-\frac{m_1m_2^2}{m}+5m_1m_2-\frac{2m_1^3}{m}\right)\frac{1}{r}-2E\left(\frac{m_1m_2^2}{2m^2}+\frac{m_1^3}{m^2}\right)-\frac{3}{2}\frac{m_1m_2^2}{m^2}\frac{h^2}{r^2}\right\}$$

$$+\frac{4m_1^2m_2}{m^2}\Sigma'\frac{(x_\mu v_\alpha-x_\alpha v_\mu)\,v_\alpha}{r^3}+\left(\frac{4m_1^3}{m^2}+\frac{3m_1^2m_2}{m^2}+\frac{3m_1m_2^2}{m^2}\right)\frac{v_\mu}{r^2}\frac{dr}{dt}. \qquad (7{\cdot}2)$$

Similarly we obtain for the acceleration of the other particle

$$\frac{d^2x_{\mu 1}}{dt^2}-\frac{m_1x_\mu}{r^3}$$

$$=-\frac{x_\mu}{r^3}\left\{\left(4m_2^2-\frac{m_1^2m_2}{m}+5m_1m_2-\frac{2m_2^3}{m}\right)\frac{1}{r}-2E\left(\frac{m_1^2m_2}{2m^2}+\frac{m_2^3}{m^2}\right)-\frac{3}{2}\frac{m_1^2m_2}{m^2}\frac{h^2}{r^2}\right\}$$

$$-\frac{4m_1m_2^2}{m}\Sigma'\frac{(x_\mu v_\alpha-x_\alpha v_\mu)\,v_\alpha}{r^3}-\left(\frac{4m_2^3}{m^2}+\frac{3m_1m_2^2}{m^2}+\frac{3m_1^2m_2}{m^2}\right)\frac{v_\mu}{r^2}\frac{dr}{dt}. \tag{7·3}$$

8. The co-ordinates $\mathbf{x}_\mu$ of the centre of mass are

$$\mathbf{x}_\mu=(m_2x_{\mu 2}+m_1x_{\mu 1})/m,$$

so that
$$\frac{d^2\mathbf{x}_\mu}{dt^2}=\frac{m_2}{m}\frac{d^2x_{\mu 2}}{dt^2}+\frac{m_1}{m}\frac{d^2x_{\mu 1}}{dt^2}. \tag{8·1}$$

Substituting from (7·2) and (7·3), we obtain

$$\frac{d^2\mathbf{x}_\mu}{dt^2}=\frac{m_1m_2(m_1-m_2)}{m^2}\left\{\frac{x_\mu}{r^3}\left(-\frac{2m}{r}-E+\frac{3}{2}\frac{h^2}{r^2}\right)+\frac{v_\mu}{r^2}\frac{dr}{dt}\right\}. \tag{8·2}$$

It is possible at this stage to compare our result with Levi-Civita's (1937*b*). Translating his equation (15) into our notation, it agrees with (8·2) except that he has $-3m/r$ instead of $-2m/r$ in the inner bracket. The secular acceleration found by him is attributable to this difference.

We now take the x_1 axis in the direction of periastron of the relative orbit. Then

$$\frac{1}{r}=\frac{1+e\cos\theta}{a(1-e^2)},\quad x_1=r\cos\theta,\quad x_2=r\sin\theta.$$

Introducing the true anomaly θ in (8·2) by the relation

$$r^2d\theta/dt=h=\sqrt{\{ma(1-e^2)\}},$$

we have
$$\frac{d}{d\theta}\frac{d\mathbf{x}_\mu}{dt}=\frac{m_1m_2(m_1-m_2)}{m^2h}\left\{\frac{x_\mu}{r}\left(-\frac{2m}{r}-E+\frac{3}{2}\frac{h^2}{r^2}\right)+v_\mu\frac{dr}{dt}\right\}.$$

Hence

$$\frac{d}{d\theta}\left(\frac{d\mathbf{x}_1}{dt}\right)=\frac{m_1m_2(m_1-m_2)}{m^2h}$$

$$\times\left\{\cos\theta\left(-\frac{2m(1+e\cos\theta)}{a(1-e^2)}-E+\frac{3}{2}\frac{m(1+e\cos\theta)^2}{a(1-e^2)}\right)-\frac{me\sin^2\theta}{a(1-e^2)}\right\}. \tag{8·3}$$

We have to pick out the secular terms on the right-hand side of (8·3). They are

$$\frac{m_1 m_2(m_1-m_2)}{mha(1-e^2)}\,e\{-1+\tfrac{3}{2}-\tfrac{1}{2}\}, \tag{8·4}$$

which is zero. Thus $\frac{d}{d\theta}\left(\frac{d\mathbf{x}_1}{dt}\right)$ is a purely periodic function of θ, so that the velocity $d\mathbf{x}_1/dt$ is a constant plus a purely periodic function. There is therefore no secular acceleration.

9. The existence of a periodic variation of the velocity of the centre of mass of a binary system illustrates our remark in § 4 on the non-vanishing of $\Sigma m_i \partial^2 r_i/\partial t^2$. It only means that the elementary geometrical definition of the centre of mass is too crude to furnish a dynamically fixed point undisturbed by the relative motion.

Since the present calculation, like those of de Sitter and Levi-Civita, is carried only to the second order, it does not exclude a secular acceleration of the third order. Since a secular acceleration implies radiation of momentum by gravitational waves, it is interesting to notice that in a comparable calculation of the radiation of energy by gravitational waves the effect found was of the third order. The radiation of gravitational energy from a double star system has not yet been calculated to the third order; but for a spinning rod a third order radiation (i.e. involving v^6) was found by Einstein (Eddington 1924, p. 248). The investigation clearly establishes that no second order radiation of energy occurs; but there is some doubt whether the positive result for the third order radiation can be relied on, owing to uncertainty in the treatment of cohesive forces (Eddington 1924, p. 251).

It is desirable to show that the secular acceleration here computed agrees with that which an observer would measure. The co-ordinates have been chosen so that at great distances from the system they approximate to Galilean co-ordinates. We can therefore surround the system with a sphere beyond which the deviation is negligible. The observer is outside the sphere and employs the Galilean co-ordinates. The acceleration will be determined from the observed change of spectral shift, which includes Doppler shift, Einstein shift, and change of length of the light-track due to the fluctuating metric. The change of Doppler shift, which corresponds to the accelerations here computed, is therefore not directly observable; nor can we, without arbitrariness, apply theoretical corrections to eliminate the other sources of shift, since these depend on the particular co-ordinate system adopted within the sphere. The periodic acceleration given by (8·4) has therefore no absolute or observable significance; but it would be possible, with some labour, to compute an observable quantity, namely the periodic variation of the total spectral shift as seen by a distant observer in a specified direction. The computation is, however, of no great interest since the effect is far below the limits of practical detection.

No such difficulty arises in observing the secular acceleration. Since the relative motion is found to be stable (§ 10), the Einstein shift and light-track correction fluctuate between narrow limits, and any secular change of the observed spectral shift must be shown in our equations as a secular change of velocity.

10. The relative orbit can also be determined from (7·2) and (7·3). The relative acceleration

$$\frac{d^2x_\mu}{dt^2}=\frac{d^2x_{\mu2}}{dt^2}-\frac{d^2x_{\mu1}}{dt^2}$$

is obtained by subtracting (7·3) from (7·2). It is unnecessary to give the details of a rather long calculation, since we confirm Levi-Civita's calculation of the potential function (1937*b*, p. 229, equation (I)) and agree with the results he derives.

References

Droste, J. 1916 *Proc. Acad. Sci. Amst.* **19**, 447–55.
Eddington, A. S. 1916 *Mon. Not. R. Astr. Soc.* **76**, 525–8.
— 1924 "Mathematical Theory of Relativity", 2nd ed. Camb. Univ. Press.
Levi-Civita, T. 1937*a* *Amer. J. Math.* **59**, 9–22.
— 1937*b* *Amer. J. Math.* **59**, 225–34.
Sitter, W. de 1916 *Mon. Not. R. Astr. Soc.* **77**, 155–84.
Tolman, R. C. 1930 *Phys. Rev.* **35**, 875–95.
Whittaker, E. T. 1935 *Proc. Roy. Soc.* A, **149**, 384–95.

Measurements with a half-pitot tube

By G. I. Taylor, F.R.S.

(*Received* 31 *March* 1938)

The use of a half-pitot tube for measuring the surface friction of a fluid flowing past a smooth surface was introduced by the late Sir Thomas Stanton (1920). The sketch, fig. 1, shows how the method works. In fig. 1, A is the half-pitot. This is a pitot tube set facing the fluid, one wall of which could, in Stanton's apparatus, be raised or lowered. The other wall is the smooth surface over which friction measurements are to be made. The difference in pressure between the fluid in the half-pitot tube and fluid in a tube connected to a hole (C, fig. 1) in the surface is measured by means of a manometer M. To calibrate the instrument it was set up in a pipe through which fluid was flowing at a speed below the critical velocity. Under these conditions the distribution of velocity is parabolic across the section of the tube, and the tangential stress at the fluid surface is known.

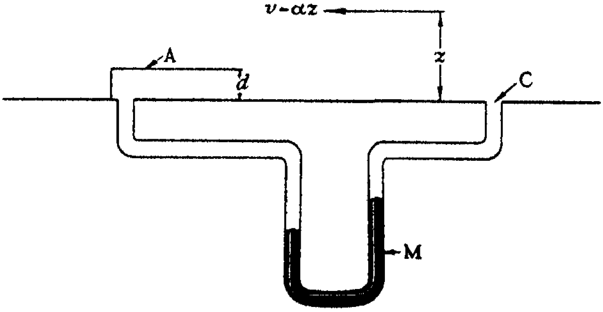


Fig. 1

Stanton expressed his results in terms of an "effective distance" d'. This is the distance from the wall at which the velocity, v, is such that the pressure in the half-pitot, p, is equal to $\frac{1}{2}\rho v^2$. When d, the opening of the pitot, was large it was found that d'/d was approximately $\frac{1}{2}$. On the other hand, it was found that when d is very small d'/d increases. Stanton's measurements of corresponding values of d' and d seem to indicate that d' tends to a finite limit as d tends to zero, so that d'/d would, in that case, tend to become infinite when d is very small. This conclusion cannot, however, be regarded as an experimentally established fact because the highest value of d'/d obtained in Stanton's experiments was only 2·2.

The idea that the pressure, p, in a half-pitot is likely to tend to a limiting finite value as d tends to zero was put forward on theoretical grounds by the present writer during a discussion of Stanton's paper. If α represents the rate of shear of the fluid at the solid surface through which the half-pitot projects, the tangential stress $f = \mu\alpha$, μ being the viscosity. On general dynamical grounds it was thought that if the Reynolds number $d^2\alpha/\nu$ is very small the flow in the neighbourhood of the half-pitot must be determined entirely by the balance of the viscous forces, the inertia forces being negligible. In this case the pressure in the half-pitot depends only on μ and α, and is independent of d. The limiting pressure, p_1, in a very small half-pitot tube must in fact be proportional to the tangential stress $\mu\alpha$.

Let

$$p_1 = kf = k\mu\alpha. \tag{1}$$

The constant k might be determined theoretically. In the work described below k was measured experimentally and found to be approximately equal to 1·2. It is hoped that the problem of determining k theoretically will attract the attention of some mathematician.

APPARATUS

In order that a low value of $d^2\alpha/\nu$ might be attained the working fluid was glycerine. The viscosity of the glycerine used was measured and found to be $\mu = 8{\cdot}2$ c.g.s. at 18° C. Its density was 1·20 so that $\nu = 6{\cdot}83$. The annular space between two concentric cylinders 2 and 3 in. diameter was filled with glycerine. The inner cylinder was rotated at constant speed by a motor. The half-pitot consisted of a very thin rectangular brass plate set at a distance of 1 mm. from the surface of the outer cylinder by two sides and a back wall to which it was soldered. The fourth side of the box formed in this way was open and faced the stream of glycerine. The brass plate forming the top of the half-pitot was 5 mm. deep in the direction of fluid flow × 10 mm. in the transverse direction. It was slightly curved so as to form part of a cylinder concentric with the outer cylinder. Tubes were lead from the half-pitot and from a small round hole cut in the outer cylinder to two sides of a mercury manometer. The level of the mercury was observed by means of a reading microscope. The general arrangement is shown in fig. 2.

The manometer gives the difference in pressure between the fluid in the half-pitot and that in a hole in the solid surface through which the half-pitot projects. If the pressure in this hole is assumed to be the same as the static pressure in the fluid, the manometer gives p, the pressure in the half-pitot.

In making observations the rotational speed of the inner cylinder was maintained constant long enough for the reading of the manometer to become steady. The number n of revolutions per second and the reading x of the microscope were observed.

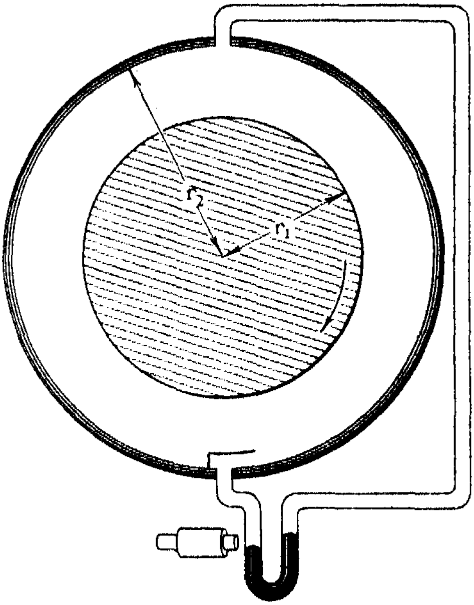


FIG. 2

RESULTS

The results are given in Table I and are shown in fig. 3, where x and n are plotted in a diagram. It will be seen that in the range covered x is proportional to n. The dotted line shown in fig. 3 represents the relation

$$x/n = 0{\cdot}72. \tag{2}$$

Calculation of k

Calibration of the manometer and reading microscope gave $p = 140x$ dynes so that the experimental result (2) is

$$p/n = 140 \times 0{\cdot}72 = 101 \text{ c.g.s.} \tag{3}$$

If r_1, r_2 are the radii of the inner and outer cylinders the velocity of the fluid at radius r is

$$v = \frac{2\pi n r_1^2}{r_2^2 - r_1^2}\left(r - \frac{r_2^2}{r}\right).$$

The tangential stress, f, at the outer cylinder is

$$f = \mu\alpha = \mu r_2 \frac{d}{dr}\left(\frac{v}{r}\right)_{r=r_2} = 4\pi\mu n r_1^2(r_2^2 - r_1^2)^{-1}.$$

TABLE I

n rev./sec.	x divisions	$d^3\alpha/\nu$	d'/d
2·35	1·6	0·035	8·3
3·25	2·4	0·048	7·1
3·5	2·4	0·051	6·9
3·7	2·7	0·055	6·7
4·15	3·0	0·061	6·3
5·15	3·7	0·076	5·7
6·30	4·5	0·093	5·1

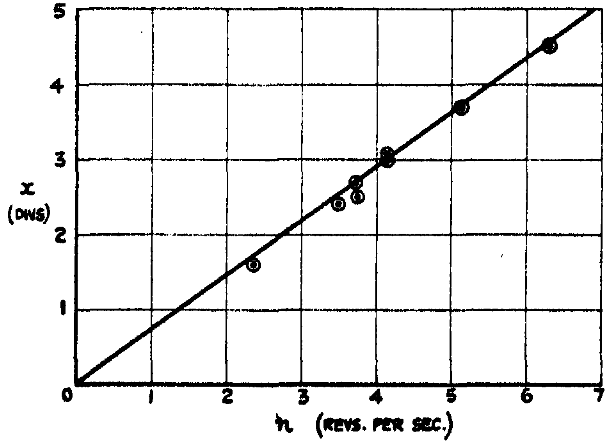


FIG. 3

In the present case $r_1 = 1·0$ in., $r_2 = 1·5$ in., so that $r_1^2(r_2^2 - r_1^2)^{-1} = 4/5$ and

$$f = 3·2\pi n\mu = 10·06 n\mu. \tag{4}$$

Hence from (3) (4) $$k = p/f = \frac{101n}{10·06n\mu} = \frac{10·0}{\mu}. \tag{5}$$

The measured value of μ at 18° C. was 8·2, so that the final result of these measurements is

$$p/f = k = 1·2. \tag{6}$$

"Effective" opening of half-pitot

The effective opening of the half-pitot tube is the distance d' at which $\frac{1}{2}\rho v^2 = p = k\mu$, and since $v = \alpha d'$

$$d' = \sqrt{\frac{2k\nu}{\alpha}}, \tag{7}$$

or

$$\frac{d'}{d} = \sqrt{\frac{2k\nu}{\alpha d^2}}. \tag{8}$$

Since αd is the velocity of the undisturbed fluid at the level of the outside of the half-pitot $\alpha d^2/\nu$ is the Reynolds number appropriate to the half-pitot. The values of $\alpha d^2/\nu$ and d'/d corresponding with the present measurements are given in cols. 3 and 4 of Table I.

Comparison with Stanton's measurements

Stanton measured the pressure in a half-pitot projecting from the wall of a pipe 0·269 cm. diameter through which air was flowing. These measurements were made with a series of values of the pitot opening d and at three mean air speeds, 955, 570 and 370 cm./sec. A selection of Stanton's measurements* of d and d' are given in Table II.

TABLE II. STANTON'S MEASUREMENTS IN 0·269 CM. PIPE.
MEAN SPEED 955 CM./SEC., $\alpha = 2{\cdot}84 \times 10^4$

d (mm.)	0·0254	0·0508	0·0762	0·1016	0·1524	0·2032	0·3556
d' (mm.)	0·0559	0·0737	0·0813	0·0940	0·1143	0·1372	0·2006
d'/d	2·20	1·45	1·06	0·925	0·760	0·59	0·564
$d^2\alpha/\nu$	1·31	5·2	10·8	20·9	47·0	83·5	256·0

To compare Stanton's results among themselves and also with the present results it is necessary to reduce them to a non-dimensional form taking account of dynamical similarity. For this reason the corresponding values of $d^2\alpha/\nu$ and d'/d have been calculated. One set of these is given in Table II. This set and also the other two sets of measurements are shown on a logarithmic scale in fig. 4. Since the flow was laminar α, the shear at the surface of the tube, was given by the expression

$$\alpha = 8\left(\frac{\text{mean velocity}}{\text{diameter of tube}}\right).$$

It will be seen in fig. 4 that Stanton's observations are fairly consistent among themselves when plotted on a non-dimensional basis. They are also consistent with the present observations in the sense that the curve through Stanton's points can be drawn tangential to the straight line through the points representing the present results.

---

* Taken from Table VI, Stanton (1920).

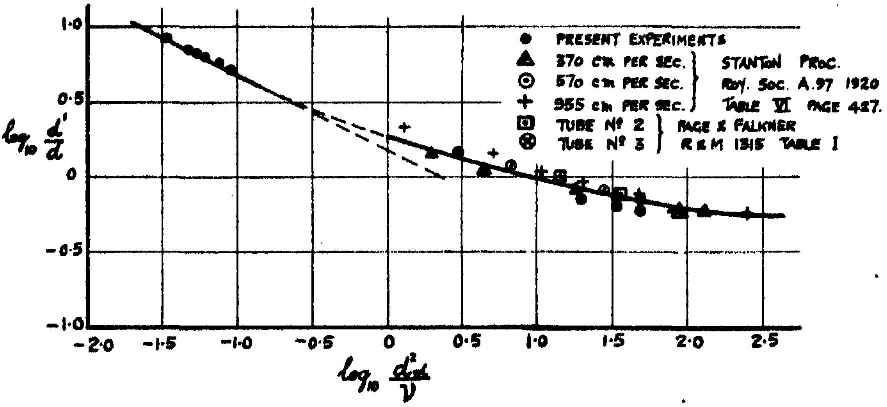


FIG. 4

Some observations by Fage and Falkner (1930) are also plotted on the same diagram. These observations were made with a flat channel instead of a round pipe. In this case the formula for finding α is

$$\alpha = 6\left(\frac{\text{mean velocity}}{\text{distance between walls}}\right).$$

REFERENCES

Fage and Falkner 1930 *Rep. Memor. Aero. Res. Comm., Lond.*, No. 1315.
Stanton 1920 *Proc. Roy. Soc.* A, **97**, 413.

The energy loss of penetrating cosmic-ray particles in copper

By J. G. Wilson

(*Communicated by P. M. S. Blackett, F.R.S.—Received* 10 *February* 1938)

[Plate 16]

1. Introduction

Absorption measurements for cosmic-ray particles by the cloud chamber method have been made in heavy metals by Anderson and Neddermeyer (1936), Crussard and Leprince Ringuet (1937), Blackett and Wilson (1937) and by Neddermeyer and Anderson (1937). Anderson and Neddermeyer have confined their attention to particles of low energy and give (1937) reasons for supposing that particles of differing absorption properties exist in this region ($E < 4 \times 10^8$ e-volts). This result is only partially confirmed by more recent work (Blackett 1938) which shows that for $E < 2 \times 10^8$ e-volts the mean energy loss of all particles and the distribution of losses among the particles are in accord with the predicted behaviour of electrons, and that there is no appreciable number of particles at this energy with non-electronic absorption properties. It is only for energies above 2×10^8 e-volts that two different kinds of particle are found. In this region the very few absorbable particles are probably normal electrons, while the penetrating particles differ in some way from normal electrons but apparently become indistinguishable from them when the energy falls below about 2×10^8 e-volts.

Measurements of the absorption of the particles in the region where penetrating particles predominate have been made by Blackett and Wilson over an extended range, while Crussard and Leprince Ringuet have measured the energy loss of particles of a mean energy about 8×10^8 e-volts and report a mean loss which is in accord with the measurements of Blackett and Wilson. Upon the nature of these penetrating particles there is not at present complete agreement. It is now generally believed that the particles are more massive than electrons and that they radiate according to the classical (Bethe-Heitler) formula appropriate to their mass.

The present measurements of the energy loss in copper cover almost the same energy range as those already reported for lead. Since the energy loss is much smaller than in lead, accuracy of measurement and freedom from systematic error in the results are of great importance. Accordingly, detailed consideration is given to the treatment of a set of data consisting of the photographs of tracks in a magnetic field.

2. The experimental method

The elimination of the large track distortions which have usually been associated with the use of metal plates in a cloud chamber has already been described (Blackett and Wilson 1937). The conditions which were then found satisfactory for measurements with a plate 1 cm. thick are also effective when a thicker (2 cm.) plate is used. The freedom from distortion as measured by the highest detectable energy is somewhat lower in the present series, and this is probably the direct result of the distortion of the gas motion by so large an obstacle.

The track photographs were taken by means of an aluminized glass mirror placed at an angle of 45° to the axis of the chamber, covering a strip of the chamber 10 cm. wide by 24 cm. The measurable part of the tracks extended right up to the metal plate at one end and almost to the edge of the metal piston at the other, a length of about 10 cm. on either side of the plate.

The purely optical distortion of track images was quite large owing to the fact that a pair of T.T.H. $f/2$ lenses of 50 mm. focus of standard type were used in place of the specially corrected 35 mm. lenses employed in the previous work. This large distortion, which is a function of the perpendicular distance of the track from the optical axis of the camera, was a drawback which probably outweighed the advantages attending the use of longer focus lenses. It was found convenient to refer the distortion arising from actual deformation of a track in the chamber to the horizontal coordinates of its midpoint above and below the plate. These two distortions were therefore separated, and the appropriate corrections to the measured track curvatures were applied separately. The optical distortions over the whole chamber were measured from photographs of a series of fine stretched wires placed inside the chamber. Measurements of a series of tracks photographed in zero magnetic field and interspaced in time with the main series of photographs, in conjunction with the optical distortions already determined, gave the values of the residual chamber distortions.

The probable error of the corrected curvature of the tracks forming the main series of photographs in a field of 10,000 gauss is represented by a maximum detectable energy of about $1{\cdot}2 \times 10^{10}$ e-volts; the error of tracks of low energy is rather greater. The energies quoted throughout the paper are those appropriate to a particle of electronic mass.

3. DETERMINATION OF THE RELATIVE ENERGY LOSS

In the case of copper, electrons up to an energy of about 5×10^8 e-volts suffer a loss of energy due to ionization which is appreciable in comparison with that due to radiative processes, and it is thus convenient to define the relative energy loss, R, so as to exclude the ionization loss. Then for all energies the smallest possible value of R is zero, and the whole observed loss can be compared directly with predicted values for the radiation loss. R is therefore defined for a thickness t of absorber by

$$R = \frac{1}{t}\frac{(E_1 - E_2 - tE_i)}{\frac{1}{2}(E_1 + E_2)}, \tag{1}$$

where E_i is the energy loss per cm. due to ionization; in this form, the maximum value of R, corresponding to the complete stoppage of the particle decreases at low energies.

The individual measurements are grouped in suitable energy ranges, and the mean energy, mean relative energy loss and the probable error of the energy loss for each range are determined. The grouping may be carried out either in terms of the measured energy of the incident particles, or of the mean measured energy in the two halves of the chamber, but for practical purposes, owing to the considerations referred to in the following paragraphs, the factors determining the choice of energy grouping are found to differ in the regions of high and of low energy.

The notable feature of the energy-loss spectrum in the low-energy region, both in the previous observations for lead and in the present series of measurements for copper, is the rapid transition from high absorption below a certain energy to very much smaller values at greater energies. In a series of measurements in which many particles in the region of high absorption lose a large fraction of their energy, the grouping of tracks according to their mean energy will tend to make the transition appear much more gradual than is in fact the case; for a particle which enters the chamber near the upper energy limit of high absorption and which almost stops is referred to a mean energy group at half this limiting energy, while smaller losses of particles of the same incident group are referred to markedly higher energies. The grouping of particles according to the incident energy is not open to this objection.

In the present series of photographs, measurements at high energies, where the absorption is small, were required, and accordingly a thick plate was used. Consequently, the low-energy particles suffered large losses, and so the tracks were grouped according to their initial energy. The group of highest energy selected in this manner extends to an incident energy of 8×10^8 e-volts.

In the high-energy region, the observed mean relative energy loss is small compared with the spread of individual values due to the probable error of measurement, and a method of grouping must be used in which this probable error does not lead to systematic error in the mean value. Such an error does in fact occur when the tracks are grouped according to their initial energy. For then the mean relative energy loss of tracks for which the initial energy has been underestimated will fall below the true value, and, moreover, the tracks will be incorporated in a group of lower initial energy. Equally, tracks of which the initial energy has been overestimated will give rise to unduly high values of the relative energy loss, and will be included in a group of higher energy. The precise effect upon the observed energy-loss spectrum depends upon the maximum detectable energy and upon the actual incident energy spectrum. In the present case the effect is to give abnormally low values of R at energies of about 10^9 e-volts, and higher values at greater energies.

For energies greater than about 5×10^8 e-volts the observed relative energy loss is both small and varies slowly. Under these conditions the grouping of observations according to mean energy on both sides of the plate is appropriate for two reasons; the probable error of the energy adopted is reduced by a factor $1/\sqrt{2}$, and the systematic error outlined above is very much reduced.

In the present series of measurements the tracks have therefore been grouped according to incident energy up to 8×10^8 e-volts and from 5×10^8 e-volts upwards according to mean energy. The two methods thus overlap in a region where neither is open to serious objection, and here the results are in agreement.

4. THE DISTRIBUTION OF ENERGY LOSS FOR CLASSICAL ELECTRONS

The distribution of energy losses for the approximate theoretical intensity distribution of emitted radiation has been treated by Heitler (1936, p. 225; cf. Bhabha and Heitler 1937, p. 436). The approximation is good for $E > 25 \times 10^6$ e-volts. Then considering only radiative losses, the probability that a particle after passing through l cm. of material has an energy e^{-y} of its initial energy is given by

$$w(y)\,dy = \frac{1}{\Gamma(l/\lambda_0)}\,e^{-y}\,y^{l/\lambda_0 - 1}\,dy,$$

where $1/\lambda_0 \fallingdotseq 18NZ^2r_0^2\alpha$, and where $\Gamma(l/\lambda_0)$ is a normalizing factor. For lead, $\lambda_0 = 0{\cdot}40$ cm., and for copper $\lambda_0 = 1{\cdot}3$ cm. Then from (1)

$$R = \frac{2}{l}\left(\frac{1-e^{-y}}{1+e^{-y}}\right),$$

so we have
$$e^{-y} = \frac{1-Rl/2}{1+Rl/2} = \frac{1-\alpha}{1+\alpha},$$

where $\alpha = Rl/2$.

Hence
$$e^{-y}\,dy = 2\frac{d\alpha}{(1+\alpha)^2},$$

and so
$$w(\alpha)\,d\alpha = 2\frac{d\alpha}{\Gamma(l/\lambda_0)}\frac{1}{(1+\alpha)^2}\left[\log\left(\frac{1+\alpha}{1-\alpha}\right)\right]^{l/\lambda_0-1}. \tag{2}$$

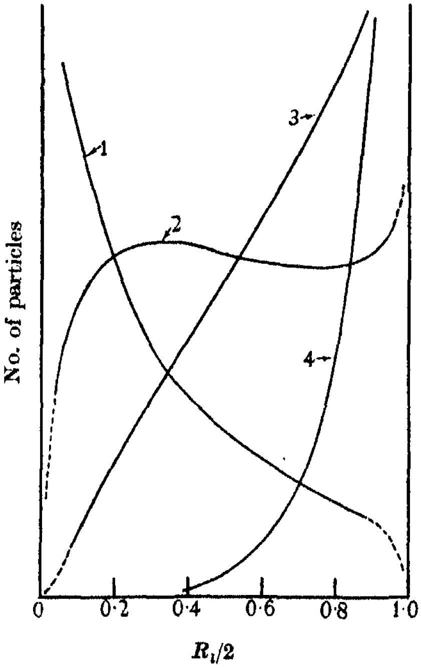


FIG. 1. Distribution of energy losses of electrons due to radiative collisions. The unit of thickness, λ_0, is that defined by Bhabha and Heitler (1937, p. 438). Curve 1, $l = 0{\cdot}8\lambda_0 = 0{\cdot}33$ cm. lead. Curve 2, $l = 1{\cdot}6\lambda_0 = 0{\cdot}64$ cm. lead or 2·0 cm. copper. Curve 3, $l = 2{\cdot}5\lambda_0 = 1{\cdot}0$ cm. lead. Curve 4, $l = 5{\cdot}0\lambda_0 = 2{\cdot}0$ cm. lead.

While the purely radiative losses considered do not provide a picture of the complete stoppage of an electron (this would of course appear in an exact treatment taking into account the loss of energy due to ionization), the electrons which are observed to be stopped have certainly suffered very large radiation losses. In fig. 1 the distribution of relative energy losses, equation (2), for a series of plate thicknesses is shown. For large relative energy losses, the effective linear path of the particle will be greater than that measured on account of the increased scattering at low energies. Thus experimental observations will tend towards curves appropriate to larger thicknesses of absorber in the region of $Rl/2 \doteqdot 1$, and an increase of the probability of very large losses is therefore to be expected.

5. THE COMPARISON OF OBSERVED ENERGY LOSS WITH THE THEORETICAL VALUE FOR ELECTRONS

The mean radiative energy loss of electrons is given (Heitler, 1936, pp. 172, 221) by

$$-\frac{1}{E}\frac{dE}{dx} = \frac{4NZ^2r_0^2}{137}\log\left(183Z^{-\frac{1}{3}}+\tfrac{2}{9}\right) \tag{3}$$

for $E \gg 137Z^{-\frac{1}{3}}mc^2$, where the symbols have their usual meanings.

We may compare the observed mean energy loss in excess of that due to ionization with this value by writing for the observed particles

$$-\frac{1}{E}\frac{dE}{dx} = k.$$

If we ignore straggling

$$E_2 = E_1e^{-kx},$$

where E_1 and E_2 are the initial and final energies of a particle passing through a thickness x of absorber.

Since from (1) $$R = (2/x)\frac{E_1-E_2-xE_t}{E_1+E_2},$$

it follows that $$k = (1/x)\log\frac{1+\bar{R}x/2}{1-\bar{R}x/2-xE_t/E_1}.$$

For $\bar{R}x \ll 1$, $k = \bar{R}+E_t/E_1$; k is a direct measure of the radiative energy loss only when $\bar{R}x \ll 1$.

6. THE MEASURED ENERGY LOSS IN COPPER

The region of high relative energy loss at low energies has not been studied in detail, since measurements were only made in a 2 cm. copper plate; to make a more thorough study of this region it would be necessary to use a much thinner plate. Throughout the experiments three counters, of which one was below the chamber, were used to control the expansions. This counter arrangement introduces a bias against large energy losses by the deflexion of very slow emergent particles away from the counter below the cloud chamber, but in the present series of tracks this effect is quite negligible. For example, under the experimental conditions, a particle of energy 3×10^9 e-volts which loses 90 % of its energy suffers a total deflexion of only about 1 cm. before reaching the counter under the cloud chamber. The counter is 10 cm. in length, and the net effect, on the chance of a particle being recorded, of deflexion away from and towards the lower counter is very much less than 10 %. For smaller losses the bias will be less. The total effect would thus be quite small for the distribution of losses of classical electrons (fig. 1, curve 2), and inappreciable for less absorbable particles.

6·1. *The curvature correction from zero field measurements*

The curvature corrections for chamber distortions were determined from forty-eight measured tracks in zero field which were interspaced in time with the main series of tracks. The residual random curvature errors of these tracks after all corrections had been applied showed negligible systematic deviation, being $+0{\cdot}01 \pm 0{\cdot}20$ m.$^{-1}$ and $+0{\cdot}02 \pm 0{\cdot}20$ m.$^{-1}$ in the upper and lower halves of the chamber. In fig. 6 (*b*) the distribution of C for these tracks is shown and the probable error of this distribution is given. The conversion factor to an energy scale was $E = (2{\cdot}64/C) \times 10^9$ e-volts, where C is the curvature in m.$^{-1}$ measured on the photographic plate; thus the maximum detectable energy was, for both halves of the chamber, $1{\cdot}3 \times 10^{10}$ e-volts.

6·2. *Measurements in the magnetic field*

The main series of photographs consists of 143 measured tracks photographed in a magnetic field of 10,000 gauss passing through a 2 cm. copper plate; for the majority of these there was no heavy material above the chamber, but for about twelve tracks 2·7 cm. of copper was over the chamber. These tracks form a complete energy spectrum under the experimental conditions, and were selected for measurement solely on considerations of photographic quality. The tracks with measured energy below $5{\cdot}5 \times 10^9$ e-volts are included directly in the energy-loss spectrum, those of higher measured energy provide a series of measurements from which any difference of behaviour between the chamber distortions in the presence and absence of the magnetic field can be deduced. Applying, in the usual way, the curvature corrections obtained from zero field tracks to correct the field tracks, such a difference will increase the probable error of measurement and introduce spurious additions to the measured relative energy losses which will be correlated with the sign of charge of the particles.

The maximum detectable energy was determined for the magnetic field photographs by two separate methods. It was found to be rather smaller than was deduced from the zero field photographs.

The first method is as follows. Nineteen tracks were observed which apparently changed sign. This property defines a group of tracks of high energy for which the real curvature is less than the error of curvature measurement. Assuming that none of the particles has suffered a measurable energy loss, the probable error of curvature measurement, ΔC, is easily shown to be approximately expressed by

$$\Delta C = \tfrac{1}{2} \times 0{\cdot}845 \frac{\Sigma | C |}{n},$$

where n is the number of tracks, each being measured above and below the plate. The measured value of $\Sigma | C |$ for these nineteen tracks gives $\Delta C = 0{\cdot}30$ m.$^{-1}$, and so a maximum detectable energy of $0{\cdot}9 \times 10^{10}$ e-volts. No case of a particle of low energy changing sign was observed.

Secondly, in the group of particles of medium energy ($2 \times 10^9 < E < 5 \times 10^9$ e-volts), the spread of the observed values of R due to errors of measurement is still much larger than the true mean value (figs. 3 and 6 (a)). It may be assumed that particles which appear to gain energy (i.e. with negative R) have all a true value of R small compared with the observed value, and that the observed distribution of apparent gains of energy is entirely due to errors of measurement. It may be deduced that the probable error of curvature measurement, ΔC, of a single track is given, from tracks in this group which appear to gain energy, by

$$\Delta C = \frac{t}{\sqrt{2}} \bar{C} \Delta R,$$

where $\bar{C}$ is the mean curvature of the group, ΔR is the probable error of these tracks (from $R = 0$), and t is the thickness of the plate. From sixteen

TABLE I. THE MAXIMUM DETECTABLE ENERGY

Method	E_m
(1) $H = 0$ tracks top half of chamber	$1{\cdot}3 \times 10^{10}$ e-volts
Lower half of chamber	$1{\cdot}3 \times 10^{10}$ e-volts
(2) Tracks changing sign of curvature	$0{\cdot}9 \times 10^{10}$ e-volts
(3) Tracks $2 < E < 5 \times 10^9$ e-volts, measured curvature decreasing	$1{\cdot}1 \times 10^{10}$ e-volts

tracks in the group which appear to gain energy $\Delta C = 0{\cdot}25$ m.$^{-1}$ and the maximum detectable energy is therefore $1{\cdot}1 \times 10^{10}$ e-volts.

The various determinations of the maximum detectable energy are consistent and are summarized in Table I.

6·3. *Difference of behaviour of the chamber with and without the magnetic field*

The systematic error of curvature which distinguishes tracks photographed in the magnetic field from those in zero field was deduced from forty-one high energy tracks with $-1{\cdot}0 < \bar{C} < +1{\cdot}0$ m.$^{-1}$ assuming that the true curvature changes of this group were small. It was only possible to ascertain the error in the measured change of curvature at the plate, as the parts of this error appropriate to the separate halves of the chamber cannot be estimated from the data available. The tracks of twenty-one positive particles gave a change of curvature of $+0{\cdot}22 \pm 0{\cdot}11$ m.$^{-1}$, and twenty negative particles $+0{\cdot}26 \pm 0{\cdot}11$ m.$^{-1}$; for all forty-one particles the change was $0{\cdot}24 \pm 0{\cdot}07$ m.$^{-1}$ (on the photographic plate).

The mean value of the relative energy loss for a group of tracks, all of about the same energy, containing equal numbers of positive and negative particles will not be appreciably affected by this source of error. When particles of either sign predominate, however, the mean relative energy loss will be displaced. Particles of a single sign which lose no energy will show a value, ΔR of the relative energy loss given by

$$\Delta R = \Delta C / C,$$

where C is the mean curvature of the tracks and ΔC is the systematic curvature error between the two halves of the chamber. When large real losses of energy occur, this relation requires a small correcting term because R is not a linear function of change of curvature. The positive and negative particles of each energy group were corrected separately and the results combined.

In the present series of measurements there is a decided predominance of positive particles, but the correction, which varies from group to group, has a mean value, as applied to ϕ/ϕ_0, of less than 0·03.

7. THE RESULTS OF THE MEASUREMENTS IN THE MAGNETIC FIELD

From the measurements of tracks in the magnetic field the relative energy loss of the particles in copper was determined up to an energy of 3×10^9 e-volts; the results of the measurements are shown in figs. 2 and 3 and in the

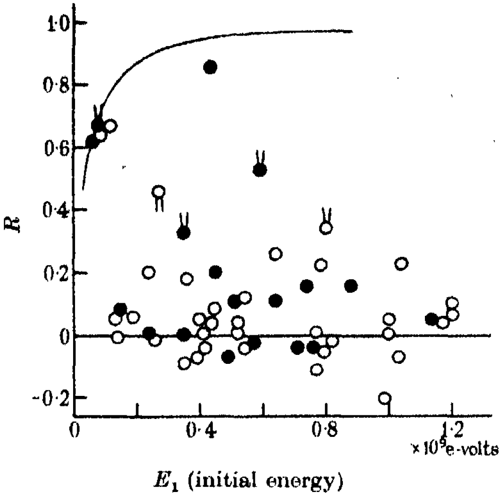


FIG. 2. The relative energy loss of tracks with $E_1 < 1{\cdot}2 \times 10^9$ e-volts. ● Negative particles. ○ Positive particles. The curved line shows the value of R corresponding to the stoppage of the particle. Points with lines above (♉) represent particles in showers.

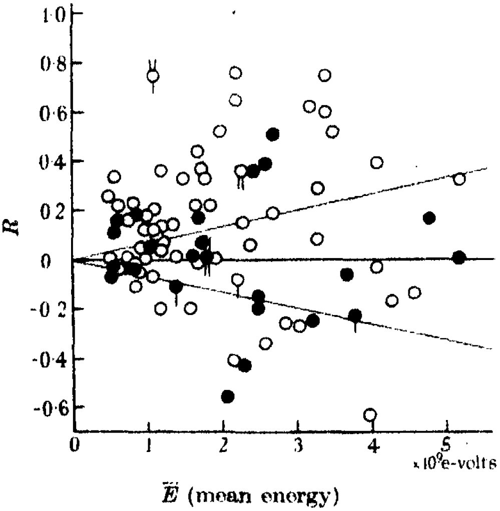


FIG. 3. Measured relative energy loss of the tracks with $\bar{E} > 0{\cdot}5 \times 10^9$ e-volts. ● Negative particles. ○ Positive particles. The thin lines show the probable error of measurement for a single track. Points with lines above represent particles in showers. Those with lines below represent particles producing secondaries, ♀ a single secondary, ⚲ more than one secondary.

first part of Table II. In fig. 2 the actual distribution of energy loss of tracks with measured initial energy less than 10^9 e-volts is shown, and in fig. 3 the distribution of energy losses of those tracks which have a measured mean energy greater than $0{\cdot}5 \times 10^9$ e-volts. In the table the tracks are grouped either by the measured incident energy or by the measured mean energy (§ 3). The method of energy grouping, the relative energy loss exclusive of ionization (equation (1)), and the cross-section for total stoppage ϕ, expressed as a fraction of the classical electron radiation cross-section, ϕ_0 (Bethe and Heitler 1934) are shown. It is inevitable that in using both methods of grouping there should be some duplication or omission of particles; in this case eleven particles appear in the table under both heads and none are omitted. The uncertainty of R and of ϕ/ϕ_0 in the table is given as the Gaussian probable error. This is directly applicable to the very high energy region, where the real loss of energy is small compared with the spread of the measured points; in the region of medium and low energies it probably somewhat overestimates the error, perhaps by as much as 50 %.

TABLE II. *The relative energy loss in copper and in lead*

Absorber and no. of tracks in group	Range of E in group (10^9 e-volts)	Mean energy of group (10^9 e-volts) and method of grouping		R	ϕ/ϕ_0
		E_1	E		
2 cm. copper 8	0–0·2	0·12	—	$0{\cdot}354 \pm 0{\cdot}098$	$0{\cdot}850 \pm 0{\cdot}235$
24	0·2–0·6	0·42	—	$0{\cdot}116 \pm 0{\cdot}041$	$0{\cdot}226 \pm 0{\cdot}077$
9	0·6–0·8	0·73	—	$0{\cdot}055 \pm 0{\cdot}031$	$0{\cdot}108 \pm 0{\cdot}061$
8	0·5–0·7	—	0·57	$0{\cdot}087 \pm 0{\cdot}039$	$0{\cdot}169 \pm 0{\cdot}076$
11	0·7–1·0	—	0·83	$0{\cdot}020 \pm 0{\cdot}023$	$0{\cdot}045 \pm 0{\cdot}051$
16	1·0–1·5	—	1·16	$0{\cdot}084 \pm 0{\cdot}020$	$0{\cdot}140 \pm 0{\cdot}034$
13	1·5–2·0	—	1·72	$0{\cdot}132 \pm 0{\cdot}033$	$0{\cdot}210 \pm 0{\cdot}053$
33	2·0–5·0	—	3·26	$0{\cdot}035 \pm 0{\cdot}042$	$0{\cdot}058 \pm 0{\cdot}070$
30	> 5·0	—	—	—	—
0·33 cm. lead 19	0–0·1	0·05	—	$2{\cdot}01 \pm 0{\cdot}27$	$1{\cdot}460 \pm 0{\cdot}196$
23	0·1–0·3	0·20	—	$0{\cdot}99 \pm 0{\cdot}14$	$0{\cdot}561 \pm 0{\cdot}079$
17	0·3–0·5	—	0·40	$0{\cdot}14 \pm 0{\cdot}08$	$0{\cdot}090 \pm 0{\cdot}049$
13	0·5–0·7	—	0·60	$0{\cdot}09 \pm 0{\cdot}14$	$0{\cdot}058 \pm 0{\cdot}090$
1 cm. lead 18	0·6–1·2	—	0·92	$0{\cdot}161 \pm 0{\cdot}040$	$0{\cdot}092 \pm 0{\cdot}027$
22	1·2–2·0	—	1·64	$0{\cdot}234 \pm 0{\cdot}045$	$0{\cdot}126 \pm 0{\cdot}024$
16	2·0–3·0	—	2·41	$0{\cdot}117 \pm 0{\cdot}071$	$0{\cdot}063 \pm 0{\cdot}038$
9	3·0–4·5	—	3·69	$0{\cdot}017 \pm 0{\cdot}059$	$0{\cdot}010 \pm 0{\cdot}035$

In the lower part of Table II the results of experiments with lead are given for purposes of comparison. In the high-energy region (1 cm. plate) the values adopted are in the main those already published (Blackett and Wilson 1937). One track which occurred on the extreme edge of the photograph and which produced a small shower has, however, been rejected, for the most energetic emergent particle may have passed out of the field of view. The measurements have also been regrouped in order to give greater uniformity between the probable errors of the group means. The measurements in the low-energy region (0·33 cm. plate) include the published data together with a considerable amount of new material (Blackett 1938).

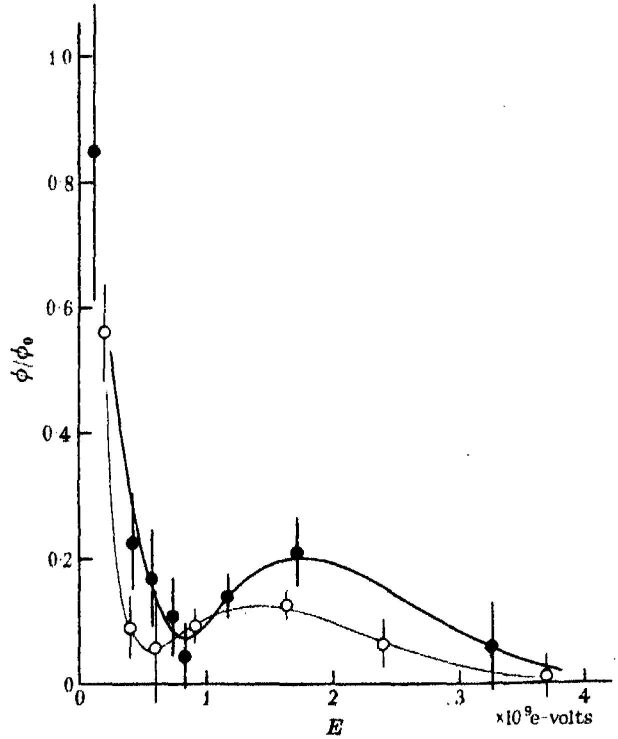


Fig. 4. Absorption cross-section (exclusive of absorption by ionization) as a function of the electron radiation cross-section. ● Copper. ○ Lead.

The important features of the revised energy-loss spectrum for lead are the very steep drop of R at energies of about $2\cdot5 \times 10^8$ e-volts to values of about 0·1 followed by a definite rise to higher values between 10^9 and 2×10^9 e-volts. The energy-loss spectrum is thus in agreement with the tentative curve of our previous paper (Blackett and Wilson 1937, p. 314, fig. 4, curve 4), and confirms the very low value of absorption then observed for an energy of 4×10^8 e-volts.

The copper measurements do not include a detailed study of the region $E \leqslant 2 \times 10^8$ e-volts, but the data applicable to this region indicate that, as in lead, a rapid diminution from the classical radiation loss of energy takes place at energies of the order of 2×10^8 e-volts. In neither case is the energy determined at which the deviation first begins. The main features of the energy loss spectrum in copper are similar to those observed in lead; the relative energy loss falls rapidly to a value of about 0·035 at about 8×10^8 e-volts and then rises again to higher values at about 2×10^9 e-volts.

8. THE ENERGY LOSS IN COPPER AND IN LEAD

The properties of practically all cosmic-ray particles at sea-level of energy greater than $2 \cdot 5 \times 10^8$ e-volts are not those of normal electrons. The present results show that the ratio ϕ/ϕ_0 falls rapidly for $E \sim 2 \cdot 5 \times 10^8$ e-volts from values of the order unity to a minimum value of less than one-tenth of this; it rises again to a maximum between 10^9 and 2×10^9 e-volts and then falls to very small values indeed for energies greater than 3×10^9 e-volts. The energy loss varies in this way for both copper and lead, and also for the mean of the measurements in both metals collected together without any distinction (i.e. assuming that ϕ/ϕ_0 is the same in copper as in lead).

The evidence of a difference of behaviour of the particles in copper and in lead is not so definite. No single point of the copper series falls more than twice its probable error from the lead curve, and so in no single case can the value of ϕ/ϕ_0 in copper be said to differ certainly from that in lead. But of the seven measured mean values of ϕ/ϕ_0 in copper at energies greater than 4×10^8 e-volts, six are greater than the corresponding values in lead, in three cases by almost twice the probable error. The single case in which ϕ/ϕ_0 in copper is less than in lead is at the minimum point of the copper spectrum, and here the trend of the minimum is indicated by the adjacent points. It must be remembered that the true errors are almost certainly rather less than those shown in fig. 4. In general, therefore the measured values of ϕ/ϕ_0 in copper are higher than those in lead. In fig. 4 the most probable forms of the ϕ/ϕ_0 spectrum in lead and copper are shown.

There seems to be no marked difference of behaviour of positive and negative particles at any point of the spectrum, and so it seems difficult to explain the excess of positive particles in the measured series as a result of anomalous absorption. It is possible that the excess of positives at sea level may be related to the excess of positives in the incident beam, as shown by the azimuthal asymmetry. Alternatively it may be an asymmetry of production in the atmosphere.

8·1. *The measured distribution of relative energy losses in copper*

The measurements in copper at low energies ($E < 3 \times 10^8$ e-volts) are insufficiently extensive to give, with any certainty, an experimental distribution of energy losses comparable with those in lead (Blackett 1938). However, the distribution of the measured energy losses is probably not inconsistent with the distribution for electrons. In figs. 5 (*a*) and 6 (*a*) the

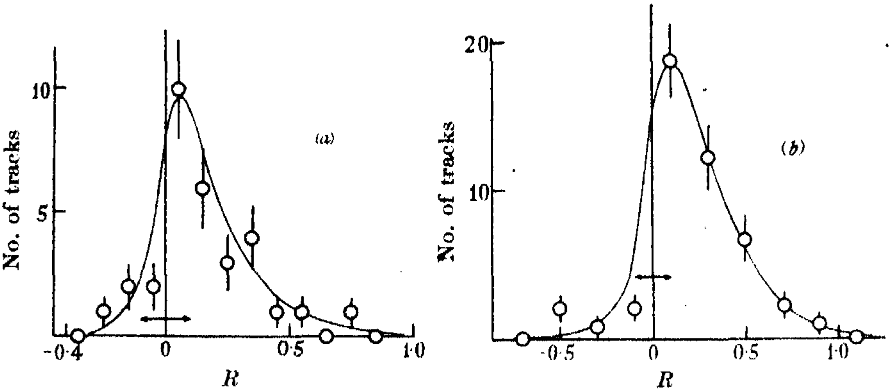


FIG. 5. Distribution of relative energy losses, (*a*) in 2 cm. copper, $10^9 < E < 2 \times 10^9$ e-volts, (*b*) in 1 cm. lead, $0{\cdot}8 \times 10^9 < E < 2{\cdot}2 \times 10^9$ e-volts. The horizontal arrows show the probable error of measurement of a single track.

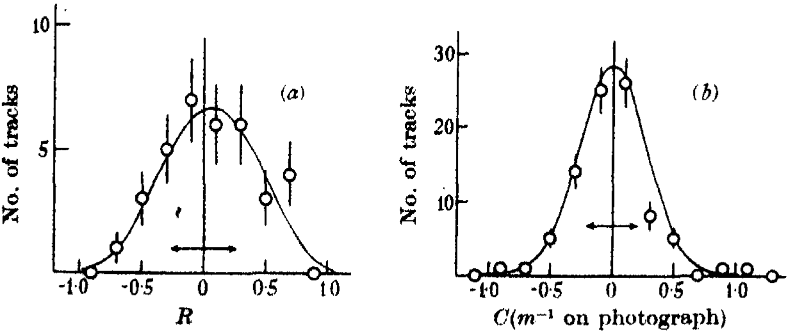


FIG. 6. Data from which the maximum detectable energy is determined. (*a*) Distribution of observed energy losses in copper, $2 \times 10^9 < E < 5{\cdot}5 \times 10^9$ e-volts. The horizontal arrow shows the probable error of measurement of a single track. (*b*) The curvature distribution of zero field tracks after all corrections have been applied. The horizontal arrow shows the deduced probable curvature error of a single track.

distribution over two ranges of higher energy is given (fig. 5 (*a*), $1{\cdot}0 \times 10^9 < E < 2{\cdot}0 \times 10^9$ e-volts; fig. 6 (*a*), $2{\cdot}0 \times 10^9 < E < 5{\cdot}5 \times 10^9$ e-volts), and these are of the same type as were found for lead at similar energies; in fig. 5 (*b*) the distribution found in lead for $0{\cdot}8 \times 10^9 < E < 2{\cdot}2 \times 10^9$ e-volts (Blackett and Wilson 1937, fig. 3 (*c*), p. 310), is repeated for purposes of comparison. From fig. 5 it is clear that the absorption at energies $\sim 1{\cdot}5 \times 10^9$ e-volts does not arise from a few particles absorbed like electrons, together with a larger number of particles losing energy only by ionization; on the contrary, the distributions are consistent with the view that practically all the particles in this part of the spectrum are of one kind. This is in agreement with other estimates of the number of electrons in this region. Blackett (1938) shows that only about 1·5 % of the particles at this energy are fully radiating electrons.

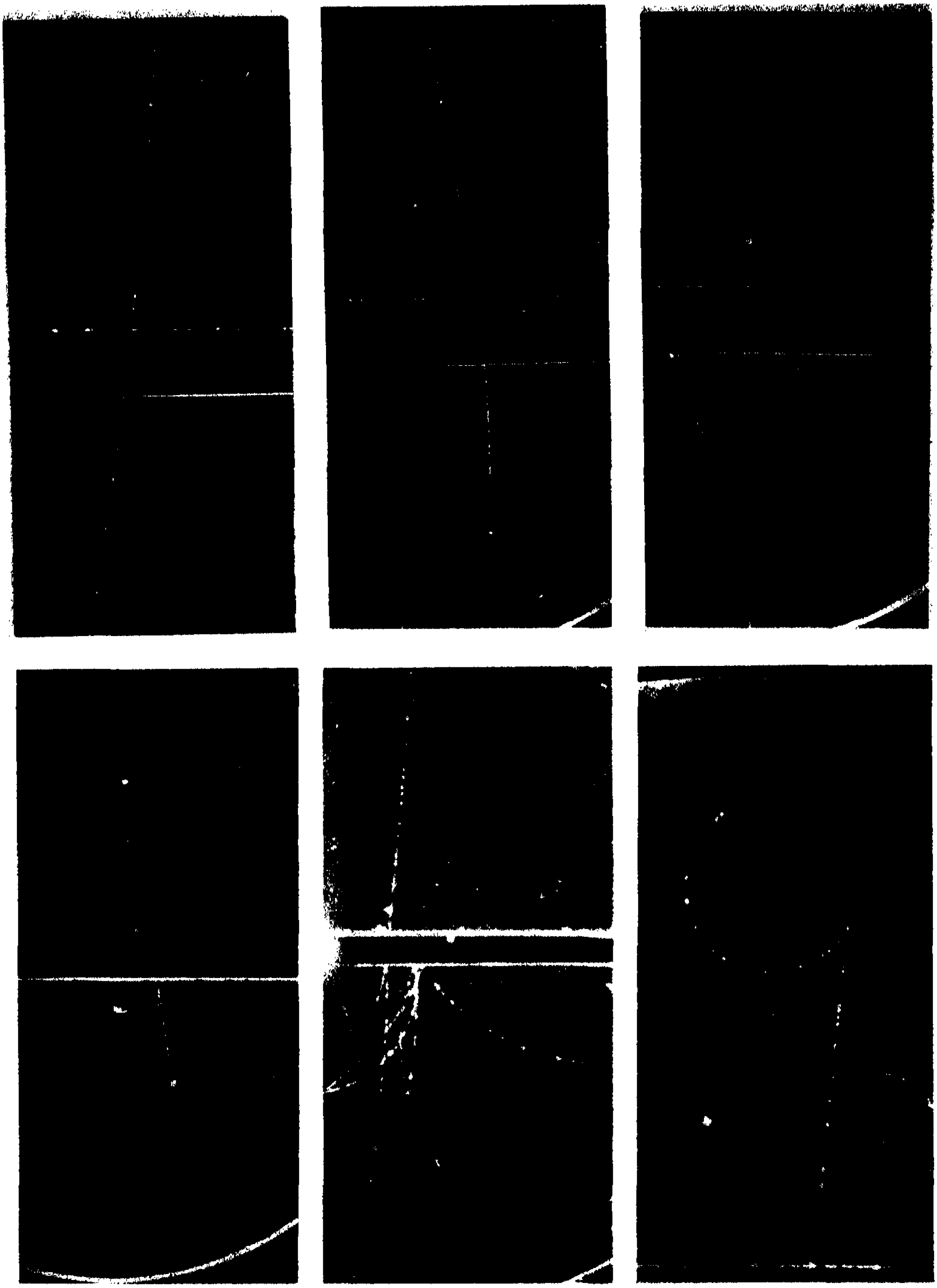
As has already been noted (§ 6·2), the distribution of measured values of R in copper for particles of very high energy, fig. 6 (*a*), is in accord with that deduced from the maximum detectable energy measured in other ways.

9. Discussion of the results

The difference in absorption between lead and copper, expressed as the ratio of the observed to the theoretical cross-section, is small, and in both metals practically all the particles with energy greater than 5×10^8 e-volts are penetrating. The absorption of the particles of low energy ($E < 5 \times 10^8$ e-volts) is considered in detail by Blackett (1938), and the present discussion relates to the more energetic, penetrating, particles.

A maximum of relative cross-section occurs for these particles in both metals at about $1{\cdot}5 \times 10^9$ e-volts. Though the observed difference in position of the two maxima (fig. 4) is within the experimental error, the maximum value of the *relative* cross-section, ϕ/ϕ_0, is most probably greater in copper than in lead. Thus the *absolute* cross-section, ϕ, in this energy region appears to increase less rapidly than as Z^2. In the region of the maximum, the mean energy loss per cm. exclusive of ionization is about 350×10^6 e-volts in lead and about 200×10^6 e-volts in copper.

It is difficult to reconcile the assumption, that this energy loss is due to radiative collisions with the subsequent emission of collision radiation in the forward direction, with the rarity of secondary particles coming from the plate, since the important quanta will have energies of the order 10^8 e-volts and less and will therefore lie on the region to which relativistic quantum mechanics certainly applies. For, on this assumption, the proportion of secondary pairs, produced by a quantum of the collision radiation, of which



7 8 9

10 11 12

Figs. 7–12

at least one of the pair electrons is observed, should be about one-third of the number of primary particles. Actually in copper, thirty-one measured tracks with $10^9 < E < 2 \times 10^9$ e-volts gave rise to two single secondaries and one pair, that is, parts of three pairs at most. Moreover, one of the secondaries and the pair were produced by the two shower particles which fall into the range, and which are most probably to be regarded as fully radiative electrons. In lead, forty-one measured tracks with $0{\cdot}8 \times 10^9 < E < 2{\cdot}2 \times 10^9$ e-volts gave rise to no secondary particles at all. There are therefore grounds for supposing that the radiation loss is very small indeed at these energies and that the observed absorption is due to some other process than the emission of collision radiation.

It has been recognized for some time that additional absorption processes may be operative at high energies, and some of these have been enumerated by Heitler (1937, p. 264). It is clear that in most cases the resulting secondaries will be canalized in the forward direction, but if the absorption is due to nuclear excitation, this will not necessarily be the case. The disintegration of the nuclei would certainly be likely in view of the large energy transfers observed, but none of the products would emerge as a rule from a metal plate.

It is not, however, possible to exclude completely the possibility that the adsorption in this energy region is due to radiative collisions for which the emission of large quanta have been suppressed. March (1937) has recently given reasons for postulating the suppression of high energy radiation, the radiated intensity becoming zero for photon energies greater than a critical value—which is defined by $\lambda = 2r_0$, that is, about 2×10^8 e-volts. In this process the probability of photon emission varies very much more rapidly than $1/\nu$, and the mean energy of the emitted photons will be very much lower than for a $1/\nu$ emission probability. The number of secondary particles observed will then be reduced on account of the increased importance of ionization energy loss and of scattering for the resulting low energy pairs. It may thus be perhaps still possible for the absence of secondaries at $E \sim 1{\cdot}5 \times 10^9$ e-volts in relation to the observed energy loss to be due to this type of radiation suppression.

In copper, the decrease of ϕ/ϕ_0 with increasing E in the region of $2{\cdot}5 \times 10^8$ e-volts is seen from fig. 4 to be distinctly less rapid than in lead, and this view is supported by the shift of the minimum value of the relative cross-section from $E \sim 5{\cdot}5 \times 10^8$ e-volts in lead to $E \sim 8{\cdot}5 \times 10^8$ e-volts in copper. The transition described by Blackett (1938) from penetrating to electron-like properties is therefore probably a function of Z. There is no indication, however, that the electron radiation formula is applicable at higher energies in copper than in lead. For the particles which are penetrating at higher energies, the radiation suppression can, in opposition to our former view, be formally represented by

$$\phi/\phi_0 = f(EZ^{-s}),$$

where $s \doteqdot \frac{1}{2}$ and not about $\frac{5}{3}$ as formerly thought. The variation of relative cross-section with Z is therefore more in accord with the form proposed by Williams (1934) than with that of Nordheim (1936).

The possibility of an explanation of the subnormal radiation loss in terms of particles of constant rest mass, but intermediate in value between that of electrons and of protons, only arises if ϕ/ϕ_0 is nearly independent of Z. It is not possible to explain the observations in terms of a single new particle losing energy by ionization and radiation only, for the lower limit of its mass would be determined by the minimum value of ϕ/ϕ_0 at about 8×10^8 e-volts (fig. 4), where the spectrum certainly includes no considerable number of protons. The higher values at $E \sim 1{\cdot}5 \times 10^9$ e-volts could then only arise from a measured electron component of about 10 % of the particles observed, and this does not appear in the distribution diagrams (fig. 5, particularly (*b*) for which $R_{\text{max}} = 2$, together with fig. 1, curve 3).

Blackett (1938) has not been able to find any penetrating particles at low energies in the unshielded spectrum, and Street and Stevenson (1937) report only one particle of intermediate mass in the equivalent of 4000 photographs under 10 cm. of lead. There is therefore some evidence that no large part of the penetrating beam remains recognizably heavy at low energies.

10. ABSORPTION OF THE PENETRATING COMPONENT IN THE ATMOSPHERE

In lead and copper, the energy loss for $E > 3{\cdot}5 \times 10^9$ e-volts is too small to measure, and the observations are consistent with a rapid drop of dE/dx as well as of $(1/E)(dE/dx)$ on the high energy side of the maximum. If the absolute cross-section at about $1{\cdot}5 \times 10^9$ e-volts varies less rapidly than Z^2, this maximum should be very pronounced for light elements. It is also probable that the position of the maximum is almost independent of Z.

Heitler (1937, also Bowen and others 1937) has shown that at the top of the atmosphere there are electron-like particles with energies up to at least 2×10^{10} e-volts, and the existence of cascade showers at sea level has given evidence for the presence there also of a few such particles. But there are too few of these particles at energies of the order $1{\cdot}5 \times 10^9$ e-volts to affect appreciably the absorption of the beam as a whole at sea-level. Then atmospheric absorption will give rise to a minimum in the energy spectrum on the high energy side of the absorption maximum (roughly at the point of inflexion of the energy loss spectrum), and since the ionization loss is comparable in magnitude, the minimum will be shallow for all thicknesses of absorber. This appears to be the most probable identification of the minimum found by Blackett (1937) in the energy spectrum at $2 \cdot 5 \times 10^9$ e-volts, for there is no indication that the initial reduction of ϕ/ϕ_0 can give rise to an absorption edge at this energy. An absorption spectrum of this type will also lead to anomalies in the variation of the vertical counting rate with altitude (Pfotzer 1936) corresponding to the initial adjustment of the incident spectrum of penetrating particles to the variable rate of absorption.

I am greatly indebted to Professor P. M. S. Blackett for affording me the opportunity of carrying out this work with the large magnet and for his constant interest and for invaluable discussions throughout. I should also like to record my appreciation of the assistance given me by Mr A. H. Chapman, who took all the photographs. The work has been made possible by a grant from the Department of Scientific and Industrial Research. The work was carried out at Birkbeck College, London.

SUMMARY

A description is given of measurements of the energy loss of cosmic ray particles in a copper plate, made by means of the cloud chamber method, and the measurements are compared with those previously made in lead.

The results show that the initial reduction of relative energy loss, $(1/E)(dE/dx)$, with increasing energy is probably rather less rapid in copper than in lead, but there is no indication that the Bethe-Heitler value for the radiation loss applies to markedly higher energies in the lighter element. For both lead and copper there is a maximum of relative energy loss at $E \sim 1 \cdot 5 \times 10^9$ e-volts. It is shown that this loss is not accompanied by the secondary particles to be expected if it were due to the emission of collision radiation. Further, the value of the relative cross-section in this region varies with the absorber less rapidly than Z^2, and the position of the maximum is almost independent of Z.

Absorption of this type will become of particular importance for elements of low atomic number, and it is suggested that the absorption with a maximum at $E \sim 1 \cdot 5 \times 10^9$ e-volts is responsible for the anomaly in the sea-level energy spectrum of cosmic rays at $E \sim 2 \cdot 5 \times 10^9$ e-volts.

REFERENCES

Anderson and Neddermeyer 1936 *Phys. Rev.* **50**, 263.
Bethe and Heitler 1934 *Proc. Roy. Soc.* A, **146**, 83.
Bhabha and Heitler 1937 *Proc. Roy. Soc.* A, **159**, 432.
Blackett 1937 *Proc. Roy. Soc.* A, **159**, 1.
— 1938 *Proc. Roy. Soc.* A, **165**, 11.
Blackett and Wilson 1937 *Proc. Roy. Soc.* A, **160**, 304.
Bowen, Millikan and Neher 1937 *Phys. Rev.* **52**, 80.
Crussard and Leprince Ringuet 1937 *C.R. Acad. Sci., Paris*, **204**, 240.
Heitler 1936 "Quantum Theory of Radiation." Oxford.
— 1937 *Proc. Roy. Soc.* A, **161**, 261.
March 1937 *Z. Phys.* **108**, 128.
Neddermeyer and Anderson 1937 *Phys. Rev.* **51**, 885.
Nordheim 1936 *Phys. Rev.* **49**, 189.
Pfotzer 1936 *Z. Phys.* **102**, 23.
Street and Stevenson 1937 *Phys. Rev.* **52**, 1003.
Williams 1934 *Phys. Rev.* **45**, 729.

DESCRIPTION OF PLATE 16

FIGS. 7, 8. Penetrating particles in the region of the relative energy loss minimum. In fig. 7 the incident energy is $(4\cdot6 \pm 0\cdot2) \times 10^8$ e-volts and the emergent energy $(4\cdot8 \pm 0\cdot2) \times 10^8$ e-volts. There is no measurable loss of energy. In fig. 8 the incident energy is $(6\cdot4 \pm 0\cdot4) \times 10^8$ e-volts, the particle loses $(2\cdot8 \pm 0\cdot5) \times 10^8$ e-volts in the copper plate (2 cm. thick). The particle is also strongly multiply scattered in the plate—through about 4·6°—this large value is not theoretically connected with the large loss of energy. It will be noticed that the tracks are undistorted right up to the surface of the plate.

FIG. 9. The primary particle is positive, and has an energy greater than 5×10^9 e-volts. The emergent particles have energies 7×10^8 e-volts (positive), 2×10^7 e-volts (negative) and 6×10^6 e-volts (negative). The primary particle is probably a normal electron, which would, on the average, produce two fast secondaries of each sign.

FIGS. 10, 11. Formation of small showers. In fig. 10, the two particles incident on the plate are almost certainly of common origin, arising above the chamber from a single photon or particle. The multiplication continues in the copper plate. The incident particle on the left is positive and has energy $(2\cdot0 \pm 0\cdot4) \times 10^9$ e-volts. It leaves the plate with $(2\cdot6 \pm 0\cdot1) \times 10^8$ e-volts. The energy of the secondary which is negative, is nearly 2×10^8 e-volts. The other incident particle is negative with an energy of $(1\cdot8 \pm 0\cdot3) \times 10^9$ e-volts; it emerges with $(1\cdot7 \pm 0\cdot3) \times 10^9$ e-volts. The energies of the pair produced are 8×10^7 e-volts (positive) and 10^7 e-volts (negative). The incident particles are probably both normal electrons, for in each case the probable number of fast secondaries is about two.

Fig. 11 shows a particle which suffers a very large loss of energy in 1 cm. of lead. The energy of the incident particle (negative) is $(7\cdot4 \pm 0\cdot4) \times 10^8$ e-volts and it leaves the plate with $(2\cdot0 \pm 0\cdot1) \times 10^8$ e-volts producing a slow negative secondary. The particle is accompanied by at least one other particle (top of photograph) and by at least one energetic photon which produces the accompanying shower. It is therefore probably an electron, although 4 or 5 secondaries would on the average be expected from an electron of this energy. The most energetic particle in the photon-produced shower has an energy of about 10^8 e-volts.

FIG. 12. Production of a fast secondary in the gas of the chamber (argon + oxygen). The primary particle is very energetic ($E \sim 10^{10}$ e-volts) and the (negative) secondary has an energy of 6×10^6 e-volts.

Magnification, figs. 7–11, × 0·45; fig. 12, × 0·65. Magnetic field, 10,000 gauss. Figs. 7–10, 2 cm. copper plate, $l = 1{\cdot}6\,\lambda_0$; fig. 11, 1 cm. lead plate, $l = 2{\cdot}5\,\lambda_0$.

On the theory of heavy electrons and nuclear forces

By H. J. BHABHA, PH.D.

Gonville and Caius College, Cambridge

(*Communicated by R. H. Fowler, F.R.S.—Received* 28 *February* 1938)

The view has been expressed by several authors, including Neddermeyer and Anderson (1937), and Street and Stevenson (1937 *a*, *b*), that the penetrating component of cosmic radiation consists largely of new particles of electronic charge and mass intermediate between those of the electron and proton, and I have shown in a recent paper (Bhabha 1938 *a*) that the shape of the transition curve for bursts and the latitude effect at sea level, which are experimentally well-established facts, also necessitate such a particle and are not compatible with a breakdown of the theory for electrons of very high energy. It has further been shown in the same paper that it does not seem to be sufficient just to postulate another particle behaving exactly like an electron of larger mass, but that the experimental evidence demands further that under certain circumstances a single heavy electron must be able to change its rest mass in the absence or presence of particles constituting ordinary matter. Indeed, the energy loss measurements of Blackett and Wilson indicate that most particles below about 2×10^8 e-volts are electrons, whereas most particles above this energy are heavy electrons, so that at about this energy there must be a large probability of a heavy electron changing or losing its identity.

Now since one may assume that charge is always conserved, it follows that there are essentially only two ways in which a single heavy electron may disappear. If, for example, it has a negative charge, it may collide with a proton and communicate its charge to it, the proton changing into a neutron, or it may turn into an ordinary electron by changing its rest mass. In either of these two processes, a certain amount of energy is liberated and the spin and statistics to be attributed to the heavy electron depend on whether the liberation of this energy is accompanied by the simultaneous liberation of some particle having a half integral spin and obeying Fermi statistics or not. Unfortunately, so far there is no experimental evidence upon this point.

Now it was shown some time ago by Yukawa (1935) that the short range and magnitude of the proton-neutron exchange interaction could be explained satisfactorily by assuming the existence of a positively charged particle without spin which could be emitted and absorbed by the change of a proton into a neutron and vice versa. The proton and neutron are here considered as two states of the same particle.* The exchange interaction thus results from the virtual exchange of such a particle between a neutron and a proton and its dependence on the separation r of the neutron and proton is of the form $e^{-\lambda r}/r$, where λ is connected with the mass M_u of the new particle by the relation $\lambda = M_u c/\hbar$. To fit the fact that nuclear forces have a range of the order of 10^{-13} cm. Yukawa then deduced that the mass of this particle had to be of the order of a hundred times the electron mass. But this is just the order of magnitude of the mass which must be attributed to the heavy electrons in cosmic radiation, as was deduced from experiment in the paper mentioned above. This value of the mass also seems to be in agreement with what has been observed by Street and Stevenson (1937*b*), and Nishina, Takeuchi and Ichimiya (1937) in Wilson chamber experiments, notwithstanding the uncertainty that attaches to these due to the extreme rarity of tracks of the required type.

It was further assumed by Yukawa that the new particle had a similar interaction with the electron and neutrino, being capable of being absorbed when an electron turned into a neutrino. On this basis it was shown by him that his theory would practically lead to the same results for the β-disintegration as the original Fermi theory. It must be mentioned, however, that in its present form it does not seem capable of explaining the asymmetrical spectrum of some of the β-decay elements. Moreover, on this theory it is possible for the new particle to disintegrate spontaneously into a positron and a neutrino, so that the possibility of identifying such particles with the heavy electrons of cosmic radiation is still open.

---

* The possible existence of a Bose particle was also suggested by Stueckelberg (1937). Kemmer (1938) has recently investigated nuclear forces on these lines, and Fröhlich and Heitler (1938) the magnetic moments of the neutron and proton.

A theory of the β-decay and nuclear forces in which spinless particles obeying Bose statistics are assumed to exist has also been proposed by Wenzel (1937 *a*, *b*). On this theory both the proton and neutron can change into both charged or uncharged particles without spin by the emission or absorption of an electron or a neutrino, whichever is consistent with the conservation of charge.* This theory seems to be so far capable of explaining the general features of the β-decay and nuclear forces. The Wenzel theory however has the great disadvantage from our point of view that the new particles must have masses near those of the proton or neutron, which in our opinion definitely excludes their being identified with heavy electrons. Nor is it possible to reformulate the theory in such a way as to allow these particles to have masses corresponding to that of the heavy electron (about a tenth of the proton mass) for this would make all nuclei unstable, inasmuch as every proton, for example, could disintegrate into a positive "Bose" particle and a neutrino.

The theory of Yukawa therefore has in our opinion the following advantages over other theories. It separates the theory of nuclear forces from that of the β-decay in such a way that the former only depends on the interaction of the proton and neutron with the heavy electron, while the latter also depends on its interaction with the electron and neutrino, so that by making the former interaction strong and the latter weak it is possible to explain the large magnitude of nuclear forces and the weakness of the β-decay. In all other theoretically self-consistent theories which have been put forward so far the two phenomena have been directly connected, so that any interaction weak enough to give a β-decay of the right magnitude has always given nuclear forces which were too small. It gives, then, an explanation of the proton-neutron exchange interaction which in non-relativistic approximation takes the form $e^{-\lambda r}/r$, where λ is connected directly with the mass of the new particle, whereas in the Fermi and Wenzel theories the force falls away inversely as a more or less high power of the distance, which it is difficult to believe is the correct description of such a fundamental thing as the force between a proton and neutron. In consequence it demands the existence of a particle of about a hundred times the electron mass, that is, of the same type as the heavy electron required in cosmic radiation. It is capable of explaining the β-decay to the same extent as the Fermi theory though not the Wenzel theory, and finally, it allows various processes which may play some role in cosmic radiation, as for example the absorption of heavy electrons by one of the particles in a nucleus, and their spontaneous disintegration into an electron and a neutrino. Since the original theory of Yukawa was not in a relativistically invariant form, it seems worth while to find a proper relativistically invariant formulation and generalization of it so that its consequences may be investigated in detail with special reference to the cosmic ray phenomenon.*

---

* The part which the collision of such particles with nuclei may play in creating showers has also been mentioned by Professor Wenzel at the Congres du Palais de la Decouverte, Paris (1937).

Following Yukawa then, the proton and neutron are considered as two states of the same particle and we assume that a proton may change into a neutron by the emission of a charged particle of mass M_u which we shall call a U-particle. It follows at once that this U-particle must have an integral spin and obey Einstein Bose statistics. A consequence of this is that the Hamiltonian for the free U-particle must be a positive definite form, since if this were not so, it would have eigenvalues corresponding to states of negative energy, which could not now be avoided by an assumption similar to that made by Dirac in the theory of the positron. Under these circumstances, as will appear below, the choice of a Hamiltonian for a free particle is very limited. A relativistic formulation of Yukawa's original theory in which the motion of the heavy electron is described by a scalar function leads just to the problem treated by Pauli and Weisskopf, but does not give the correct form of nuclear forces.† The simplest generalization is therefore to describe the motion of U-particles by four wave-functions transforming like a four-vector. This is the case that is here treated in detail, and it will appear below that the possible forms of the interaction of the heavy electron with the proton-neutron are very limited and contain only two arbitrary constants. This treatment already gives the forces we require, so that we have not investigated equations of higher tensor form.‡

It is obvious that the U-particles, being charged, can never lead to a proton-proton force except as a fourth order process, corresponding to the emission of U-particles by each of two protons and their absorption by the other. Now the proton-proton force as deduced by Breit, Condon and Present (1936) from experiments by Tuve, Heydenberg and Hafstad (1936) appears to be the same as the neutron-proton force to within about 30 %. In order to have a proton-proton interaction of the same order as the proton-neutron force, it appears necessary to introduce a neutral particle N obeying Bose statistics which may be emitted or absorbed when a proton jumps from one energy state to another. I do not believe, however, that it is yet possible to say that the introduction of a neutral particle is *necessary*, since the usual approximation methods of quantum mechanics, upon which this conclusion is based, consisting in a development in powers of those constants here which correspond to the fine-structure constant $e^2/\hbar c$ of the radiation theory, is not valid inasmuch as these constants have a value which lies between a third and a tenth, and are not small compared to unity. However, I consider the existence of such a particle as not improbable, since it corresponds to a symmetrical state of things in which all masses fall into three groups of the order of magnitude of the electron mass, the U-particle mass, and the proton mass, and that we would have positive, negative and neutral particles in each group.* It must be emphasized, however, that the introduction of such particles has nothing to do formally with the theory of U-particles which can be formulated without them, and is in addition to it. The theory of neutral Bose particles can be formulated exactly as the theory of U-particles with trivial modifications.

---

* I am very much indebted to Dr W. Heitler for a discussion on cosmic radiation in which he drew my attention to the theory of Yukawa. This discussion formed the starting point of the considerations of this paper.

† Since these calculations were completed, a new paper has appeared by Yukawa and Sakata (1937) in which just this case has been treated. As their results show, the theory gives a Heisenberg exchange force giving only a stable singlet state to the deuteron, which is contrary to observation.

‡ Taking the wave functions to be antisymmetrical in all the suffices, these reduce in essence to the cases dual to those treated by Yukawa and ourselves. An analysis of all the various cases in relation to nuclear forces has been carried out by Dr N. Kemmer, and I am indebted to him for informing me that all the other cases except the one treated here also give the wrong nuclear forces.

Finally, it may be mentioned that this theory also leads to showers of Heisenberg's type, with this difference that these showers would consist mostly of heavy electrons and some proton-neutrons, but few electrons. The existence of these showers still further invalidates the usual methods of calculation, since it shows that for high energies higher order processes are just as important as the first-order process.

1. THE EQUATIONS OF MOTION

In all problems where equations of motion are to be found for some system it is convenient to derive the equations by the variation of some relativistically invariant Lagrange function to ensure that the conservation laws should automatically hold. We shall adopt this procedure, treating all the quantities classically, and then proceed to the quantization in the next section.

---

* The theory of Wenzel also contains two new particles, one charged and one uncharged.

The simplest assumption is to suppose that the motion of the U-particles is described by some invariant wave function U. This would lead to the scalar relativistic wave equation for U which has been quantized by Pauli and Weisskopf (1934). This assumption has been made by Yukawa and Sakata (1937) in a recent paper and does not lead to the correct form for the neutron-proton interaction, as we have already mentioned in the introduction. We therefore make the next simplest assumption, namely that the motion of the U-particles is described by four wave functions, U_0, U_1, U_2, and U_3, which transform like a four-vector. In order to treat the time component on the same footing as the space components, following Proca (1936) we introduce a quantity ϵ such that $\epsilon^2 = -1$, and as usual define a quantity U_4 by

$$U_4 = \epsilon U_0.$$

But ϵ must not be confounded with i, the other root of -1 which appears in our equations, and in going over to the conjugate complex equations it must be remembered that whereas the sign of i is changed, the sign of ϵ is not.*

Now the simplest relativistically invariant Lagrange function which contains only the U_μ† and their derivatives with regard to the co-ordinates x_k and $x_4 \equiv \epsilon ct$ is, according to Proca, of the form

$$\left|\left(\frac{\partial U_\nu}{\partial x_\mu}-\frac{\partial U_\mu}{\partial x_\nu}\right)\right|^2+\text{const. }\overline{U_\mu}U_\mu.$$

The $\partial/\partial x_\mu$ here, as usual, play the part of the momentum operator p_μ, so that to generalize the above Lagrangian when a field is present we replace

$$\frac{\partial}{\partial x_\mu} \quad \text{by} \quad \frac{\partial}{\partial x_\mu}-\frac{ie}{\hbar c}\phi_\mu,$$

where ϕ_μ are the electromagnetic potentials. We therefore introduce the quantities $G_{\mu\nu}$ defined by

$$G_{\mu\nu} = \left(\frac{\partial}{\partial x_\mu}-\frac{ie}{\hbar c}\phi_\mu\right)U_\nu-\left(\frac{\partial}{\partial x_\nu}-\frac{ie}{\hbar c}\phi_\nu\right)U_\mu, \tag{1}$$

---

* This is merely based on the elementary fact that if U_0, U_1, U_2, U_3, form a complex four-vector, U_1, U_2, U_3, and ϵU_0 form the corresponding Euclidean four-vector, and the Euclidean four-vector corresponding to the conjugate complex vector $\overline{U_0}$, $\overline{U_1}$, $\overline{U_2}$, $\overline{U_3}$, is just $\overline{U_1}$, $\overline{U_2}$, $\overline{U_3}$, and $\epsilon\overline{U_0}$.

† Greek suffices run from 1 to 4, Latin suffices from 1 to 3. A summation is understood over repeated suffices.

and take as the relativistically invariant Lagrange function for the U-particles in the presence of a field (Proca 1936)

$$L^U = -\tfrac{1}{2}\hbar^2 c^2 \overline{G}_{\mu\nu} G_{\mu\nu} - M_u^2 c^4 \overline{U}_\mu U_\mu, \tag{2}$$

where M_u is the mass of the U-particles, and $\hbar$ is Planck's constant divided by 2π.

As stated in the introduction, the neutron and proton are considered as two states of the same particle described by the eigenvalues $+1$ and -1 respectively of some operator τ say. We then introduce the matrices

$$\tau_N = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}; \; \tau_P = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}; \; \tau_{NP} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}; \; \tau_{PN} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{3}$$

which satisfy the relations

$$\tau_P \tau_{PN} = \tau_{PN} \tau_N = \tau_{PN}; \; \tau_N \tau_{PN} = \tau_{PN} \tau_P = 0, \text{ etc.} \tag{3a}$$

and

$$\left.\begin{aligned} \tau_{NP}\tau_{PN} &= \tau_N, \\ \tau_{PN}\tau_{NP} &= \tau_P. \end{aligned}\right\} \tag{3b}$$

The above relations are easily remembered if it is noticed that all those products are zero in which dissimilar suffices appear in succession in two neighbouring matrices. ψ will denote the wave function for the proton-neutron, which now consists of two components belonging to the proton and neutron respectively. We assume that the motion of the free neutron or proton is given by the Dirac equation, so that each of the two components is really composed of the usual four wave functions for such a particle. The objection to the use of the Dirac equation here on the ground that it does not give a correct description of the magnetic moment of the proton or neutron is not strictly valid, since the proton or neutron possess an additional magnetic moment due to their interaction with the U-particles, exactly as in the Fermi theory. The Lagrangian for the proton-neutron is assumed to be of the form

$$L^M = -\psi^+\left[c\gamma^\mu\left(p_\mu - \frac{e}{c}\phi_\mu \tau_P\right) - ic^2(M_N\tau_N + M_P\tau_P)\right]\psi. \tag{4}$$

Here p_k is the momentum and $p_4 = \epsilon E/c$, E being the energy, and M_N, M_P are the masses of the neutron and proton respectively. Denoting by α^k and β the four Dirac matrices and writing $\alpha^4 \equiv \epsilon$, we introduce the matrices γ^μ defined by

$$\gamma^\mu \equiv -i\beta\alpha^\mu, \tag{5a}$$

and define ψ^+ by

$$\psi^+ = \epsilon\overline{\psi}\gamma^4 = i\overline{\psi}\beta. \tag{5b}$$

Then, as is well known, $\psi^+\gamma^\mu\psi$ form a Euclidian four-vector, and $\psi^+\gamma^\mu\gamma^\nu\psi$ an antisymmetrical Euclidean tensor with all components real,[*] so that

$$\overline{\psi^+\gamma^\mu\psi} = \psi^+\gamma^\mu\psi$$

and

$$\overline{\psi^+\gamma^\mu\gamma^\nu\psi} = \psi^+\gamma^\mu\gamma^\nu\psi. \tag{5c}$$

Some assumption must now be made about the interaction between the proton-neutron and U-particles. The simplest interaction that can be assumed is of the type $\psi^+\gamma^\mu(U_\mu\tau_{PN}+\overline{U}_\mu\tau_{NP})\psi$ consisting of the product of two four-vectors, as I have already indicated in a preliminary communication (1938*b*). The next simplest interaction is of the form

$$\psi^+\gamma^\mu\gamma^\nu(G_{\mu\nu}\tau_{PN}+\overline{G_{\mu\nu}}\tau_{NP})\psi$$

consisting of the products of two antisymmetrical tensors. Now, as will appear below, there is a certain symmetry in the equations for the G's and U's, and the two interactions given above are really on the same footing.[†] Consideration is therefore restricted to these two terms only, and we assume for the interaction between the U-particles and the proton-neutron a term of the form

$$L^I = -g_1\psi^+\gamma^\mu(U_\mu\tau_{PN}+U_\mu\overline{\tau_{NP})}\psi \tag{6}$$
$$-\tfrac{1}{2}g_2\psi^+\gamma^\mu\gamma^\nu(G_{\mu\nu}\tau_{PN}+\overline{G_{\mu\nu}}\tau_{NP})\psi,$$

where g_1 and g_2 are two independent constants.

A similar Lagrange function must be added for the electron neutrino, but we shall omit this until a later section for brevity. The Lagrange function for the electromagnetic field is as usual

$$L^{\text{Max}} = -\frac{1}{16\pi}F_{\mu\nu}F_{\mu\nu} = \frac{1}{8\pi}(\mathbf{E}^2-\mathbf{H}^2), \tag{7}$$

E and **H** being electric and magnetic force, where

$$E_k = F_{0k}, \quad H_{lm} = F_{lm}$$

and

$$F_{\mu\nu} = \frac{\partial}{\partial x_\mu}\phi_\nu - \frac{\partial}{\partial x_\nu}\phi_\mu. \tag{8}$$

The equations of motion for the whole system may then be derived as usual from the principle that for arbitrary variations of ψ, U_μ, ϕ_μ and their conjugates the change in the proper Lagrangian $\mathscr{L}$ for the whole system shall be zero, where

---

[*] Cf. W. Pauli (1933, p. 221). In the usual formulation without the quantity ϵ, the space components are purely real whereas the time components are purely imaginary.

[†] I am indebted to Dr N. Kemmer for drawing my attention to this fact.

$$\mathscr{L} = \iiiint dx_1\,dx_2\,dx_3\,dx_4(L^M + L^U + L^I + L^{\mathrm{Max}}). \tag{9}$$

Varying $\overline{\psi}$ gives

$$\begin{aligned}\Big[c\gamma^\mu\Big(p_\mu - \frac{e}{c}\phi_\mu \tau_P\Big) - ic^2(M_N\tau_N + M_P\tau_P)\Big]\psi \\ + [g_1\gamma^\mu(U_\mu\tau_{PN} + \overline{U_\mu}\tau_{NP}) + \tfrac{1}{2}g_2\gamma^\mu\gamma^\nu(G_{\mu\nu}\tau_{PN} + \overline{G_{\mu\nu}}\tau_{NP})]\psi = 0,\end{aligned} \tag{10a}$$

which, eliminating ϵ, reads

$$\begin{aligned}\Big[c\alpha^k\Big(p_k - \frac{e}{c}\phi_k\tau_P\Big) - \Big(E_0 - \frac{e}{c}\phi_0\tau_P\Big) + \beta c^2(M_N\tau_N + M_P\tau_P) \\ + g_1\{\alpha^k(U_k\tau_{PN} + \overline{U_k}\tau_{NP}) - U_0\tau_{PN} - \overline{U_0}\tau_{NP}\} \\ + \tfrac{1}{2}g_2\{i\beta\alpha^k\alpha^l(G_{kl}\tau_{PN} + \overline{G_{kl}}\tau_{NP}) + 2i\beta\alpha^l(G_{0l}\tau_{PN} + \overline{G_{0l}}\tau_{NP})\}\Big]\psi.\end{aligned} \tag{11}$$

Variation of ψ gives the conjugate equation

$$\begin{aligned}\psi^+\Big[c\gamma^\mu\Big(p_\mu - \frac{e}{c}\phi_\mu\tau_P\Big) - ic^2(M_N\tau_N + M_P\tau_P) \\ + g_1\gamma^\mu(U_\mu\tau_{PN} + \overline{U_\mu}\tau_{NP}) + \tfrac{1}{2}g_2\gamma^\mu\gamma^\nu(G_{\mu\nu}\tau_{PN} + \overline{G_{\mu\nu}}\tau_{NP})\Big],\end{aligned} \tag{10b}$$

where p_μ acting backwards on ψ^+ is as usual to be understood as the operator $i\hbar\partial/\partial x_\mu$ instead of $-i\hbar\partial/\partial x_\mu$.

Variation of $\overline{U_\mu}$ gives

$$\begin{aligned}\hbar^2c^2\Big(\frac{\partial}{\partial x_\mu} - \frac{ie}{\hbar c}\phi_\mu\Big)G_{\mu\nu} - M_u^2c^4U_\nu - g_1\psi^+\gamma^\nu\tau_{NP}\psi \\ + g_2\Big(\frac{\partial}{\partial x_\mu} - \frac{ie}{\hbar c}\phi_\mu\Big)\{\psi^+\gamma^\mu\gamma^\nu\tau_{NP}\psi\}_{\mu\neq\nu} = 0,\end{aligned} \tag{12a}$$

and variation with respect to U_μ gives the equations conjugate to (12), namely

$$\begin{aligned}\hbar^2c^2\Big(\frac{\partial}{\partial x_\mu} + \frac{ie}{\hbar c}\phi_\mu\Big)\overline{G_{\mu\nu}} - M_u^2c^4\overline{U_\nu} - g_1\psi^+\gamma^\nu\tau_{PN}\psi \\ + g_2\Big(\frac{\partial}{\partial x_\mu} + \frac{ie}{\hbar c}\phi_\mu\Big)\{\psi^+\gamma^\mu\gamma^\nu\tau_{PN}\psi\}_{\mu\neq\nu} = 0.\end{aligned} \tag{12b}$$

Introducing a quantity $G'_{\mu\nu}$ defined by*

$$G'_{\mu\nu} = \hbar c\, G_{\mu\nu} + \frac{g_2}{\hbar c}\psi^+\gamma^\mu\gamma^\nu\tau_{NP}\psi \tag{13}$$

and its conjugate, and using (1) we get

$$G'_{\mu\nu} = \hbar c\left\{\left(\frac{\partial}{\partial x_\mu} - \frac{ie}{\hbar c}\phi_\mu\right)U_\nu - \left(\frac{\partial}{\partial x_\nu} - \frac{ie}{\hbar c}\phi_\nu\right)U_\mu\right\} + \frac{g_2}{\hbar c}\psi^+\gamma^\mu\gamma^\nu\tau_{NP}\psi, \tag{14a}$$

and $$\overline{G'_{\mu\nu}} = \hbar c\left\{\left(\frac{\partial}{\partial x_\mu} + \frac{ie}{\hbar c}\phi_\mu\right)\overline{U}_\nu - \left(\frac{\partial}{\partial x_\nu} + \frac{ie}{\hbar c}\phi_\nu\right)\overline{U}_\mu\right\} + \frac{g_2}{\hbar c}\psi^+\gamma^\mu\gamma^\nu\tau_{PN}\psi. \tag{14b}$$

The equations (12) may now be written

$$\hbar c\left(\frac{\partial}{\partial x_\mu} - \frac{ie}{\hbar c}\phi_\mu\right)G'_{\mu\nu} - M_u^2c^4U_\nu - g_1\psi^+\gamma^\nu\tau_{NP}\psi = 0, \tag{15a}$$

$$\hbar c\left(\frac{\partial}{\partial x_\mu} + \frac{ie}{\hbar c}\phi_\mu\right)\overline{G'_{\mu\nu}} - M_u^2c^4\overline{U}_\nu - g_1\psi^+\gamma^\nu\tau_{PN}\psi = 0 \tag{15b}$$

It is now seen that there is complete symmetry between the G''s and U's, the equations (14) expressing the G''s in terms of the rotation of the U's, while the equations (15) express the U's in terms of the divergence of the G''s. These equations show us that the two interaction terms in (6) are on exactly the same footing.

Finally, varying ϕ_μ,

$$\frac{1}{4\pi}\frac{\partial}{\partial x_\nu}F_{\nu\mu} + ie(U_\nu\overline{G'_{\mu\nu}} - \overline{U}_\nu G'_{\mu\nu}) + e\psi^+\gamma^\mu\tau_P\psi = 0. \tag{16}$$

These are just the four Maxwell equations.

$$\frac{1}{4\pi}\left(\text{curl}_k\,\mathbf{H} - \frac{1}{c}\mathbf{E}_k\right) = ie(U_\nu\overline{G'_{k\nu}} - \overline{U}_\nu G'_{k\nu}) + e\overline{\psi}\alpha^k\tau_p\psi, \tag{17a}$$

and $$\frac{1}{4\pi}\text{div}\,\mathbf{E} = ie(U_k\overline{G'_{0k}} - \overline{U}_k G'_{0k}) + e\overline{\psi}\tau_P\psi. \tag{17b}$$

The second term on the right-hand side is just the usual charge current S_μ for the protons

$$S_\mu = e\overline{\psi}\alpha^\mu\tau_P\psi, \tag{18}$$

while the first term gives the expression for the current due to the U-particles

$$\sigma_\mu = ie(U_\nu\overline{G'_{\mu\nu}} - \overline{U}_\nu G'_{\mu\nu}), \tag{19}$$

---

* We make the convention that $G'_{\mu\nu}$ is identically zero when the two suffices are equal.

as has already been found by Proca in the absence of the interaction (6). Applying the operator $\partial/\partial x_\mu$ to the equation (16), the left-hand side vanishes, so that

$$\frac{\partial}{\partial x_\mu}(S_\mu+\sigma_\mu) = 0. \tag{20}$$

However, the divergence of each part separately is *not* zero, which is obviously due to the fact that charge may be transferred from the proton to the U-particle. Using the equations (12) and (10), it can easily be shown that

$$\frac{\partial \sigma_\mu}{\partial x_\mu} = \frac{ie}{\hbar c}[g_1 \psi^+ \gamma^\nu (U_\nu \tau_{PN} - \overline{U}_\nu \tau_{NP})\psi + \tfrac{1}{2} g_2 \psi^+ \gamma^\mu \gamma^\nu (G_{\mu\nu}\tau_{PN} - \overline{G}_{\mu\nu}\tau_{NP})\psi] = -\frac{\partial S_\mu}{\partial x_\mu}, \tag{21}$$

which is an explicit expression for the rate at which the U-particles are created due to the proton-neutron alone. It should be noticed that (20) completely determines the correct sign of e in (1), since with the opposite sign of e the sign of the first expression on the left of equation (21) would be changed, thus violating (20) and leading to an inconsistency.

To find the Hamiltonian corresponding to the material part of the Lagrangian (9), we seek the energy-momentum tensor $T_{\mu\nu}$, a quantity which by definition satisfies the equation

$$\frac{\partial}{\partial x_\mu} T_{\mu\rho} = F_{\rho\nu}(S_\nu+\sigma_\nu). \tag{22}$$

The method here used follows that of Proca closely. Writing

$$L = L^M + L^U + L^I.$$

and using the well-known Lagrange equations,

$$\frac{\partial L}{\partial x_\rho} = \frac{\partial}{\partial x_\mu}\left[\frac{\partial L}{\partial\left(\frac{\partial U_\nu}{\partial x_\mu}\right)}\cdot\frac{\partial U_\nu}{\partial x_\rho} + \text{conj.} + \frac{\partial L}{\partial\left(\frac{\partial \psi}{\partial x_\mu}\right)}\cdot\frac{\partial \psi}{\partial x_\rho} + \text{conj.}\right] + \frac{\partial L}{\partial \phi_\mu}\cdot\frac{\partial \phi_\mu}{\partial x_\rho}. \tag{23}$$

Then, remembering (15),

$$F_{\rho\mu}(S_\mu+\sigma_\mu) = \left(\frac{\partial \phi_\mu}{\partial x_\rho} - \frac{\partial \phi_\rho}{\partial x_\mu}\right)\frac{\partial L}{\partial \phi_\mu} = -\frac{\partial}{\partial x_\mu}\left(\phi_\rho \frac{\partial L}{\partial \phi_\mu}\right) + \frac{\partial \phi_\mu}{\partial x_\rho}\frac{\partial L}{\partial \phi_\mu},$$

and using (18)

$$F_{\rho\mu}(S_\mu+\sigma_\mu) = -\frac{\partial}{\partial x_\mu}\left[\phi_\rho\frac{\partial L}{\partial\phi_\mu}+\frac{\partial L}{\partial\left(\frac{\partial U_\nu}{\partial x_\mu}\right)}\cdot\frac{\partial U_\nu}{\partial x_\rho}\right.$$
$$\left.+\text{conj.}+\frac{\partial L}{\partial\left(\frac{\partial\psi}{\partial x_\mu}\right)}\cdot\frac{\partial\psi}{\partial x_\rho}+\text{conj.}-\delta_{\mu\rho}L\right]$$
$$=\frac{\partial}{\partial x_\mu}\left[\hbar c\overline{G'_{\mu\nu}}\left(\frac{\partial}{\partial x_\rho}-\frac{ie}{\hbar c}\phi_\rho\right)U_\nu+\text{conj.}+\psi^+\gamma^\mu c\left(p_\rho-\frac{e}{c}\phi_\rho\tau_P\right)\psi+\delta_{\mu\rho}L\right]. \tag{24}$$

Further,

$$\frac{\partial}{\partial x_\mu}\left\{-\hbar c\overline{G'_{\mu\nu}}\left(\frac{\partial}{\partial x_\nu}-\frac{ie}{\hbar c}\phi_\nu\right)U_\rho+M_u^2c^4\overline{U_\mu}U_\rho\right\}+\text{conj.}$$
$$=\frac{\partial}{\partial x_\mu}\left\{-\hbar c\frac{\partial}{\partial x_\nu}(\overline{G'_{\mu\nu}}U_\rho)-\hbar c\left(\frac{\partial}{\partial x_\nu}+\frac{ie}{\hbar c}\phi_\nu\right)\overline{G'_{\mu\nu}}U_\rho+M_u^2c^4\overline{U_\mu}U_\rho\right\}+\text{conj.}$$
$$=\frac{\partial}{\partial x_\mu}\left\{-g_1\psi^+\gamma^\mu(U_\rho\tau_{PN}+\overline{U_\rho}\tau_{NP})\psi\right\}.$$

Adding this equation to (21) we get (22) with

$$T_{\mu\rho}=\left[\psi^+\gamma^\mu c\left(p_\rho-\frac{e}{c}\phi_\rho\tau_P\right)\psi+\hbar^2c^2\overline{G_{\mu\nu}}G_{\rho\nu}+\text{conj.}\right.$$
$$+M_u^2c^4(\overline{U_\mu}U_\rho+\overline{U_\rho}U_\mu)+\delta_{\mu\rho}L$$
$$+g_1\psi^+\gamma^\mu(U_\rho\tau_{PN}+\overline{U_\rho}\tau_{NP})\psi$$
$$\left.+g_2\psi^+\gamma^\mu\gamma^\nu(G_{\rho\nu}\tau_{PN}+\overline{G_{\rho\nu}}\tau_{NP})\psi\right]. \tag{25}$$

This is the energy-momentum tensor for the material part of the system. The energy is then just

$$T_{00}=-T_{44}=\psi^+\left\{\gamma^k c\left(p_k-\frac{e}{c}\phi_k\tau_P\right)-ic^2(M_N\tau_N+M_P\tau_P)\right\}\psi$$
$$+\{\hbar^2c^2(\overline{G_{0k}}G_{0k}+\tfrac{1}{2}\overline{G_{kl}}G_{kl})+M_u^2c^4(\overline{U_0}U_0+\overline{U_k}U_k)\}$$
$$+g_1\psi^+\gamma^k(U_k\tau_{PN}+\overline{U_k}\tau_{NP})\psi$$
$$+\tfrac{1}{2}g_2\psi^+\gamma^k\gamma^l(G_{kl}\tau_{PN}+\overline{G_{kl}}\tau_{NP})\psi. \tag{26}$$

The first term represents the energy of the proton-neutron, the second term that of the U-particles, and the last term the interaction energy. The

Hamiltonian for the free U-particles is a positive definite form. One should notice particularly that here, as in the electromagnetic field, the time components U_0 and G_{0k} are absent from the interaction.

2. QUANTIZATION OF THE EQUATIONS OF MOTION

To quantize these equations one has now only to find the right commutation rules for the G's and U's. The momenta conjugate to U_k and $\overline{U_k}$ are $\overline{G_k}$ and G_k respectively, given by

$$\overline{G_k} \equiv \frac{1}{\hbar}\frac{\partial L}{\partial\left(\dfrac{\partial U_k}{\partial t}\right)} = -\overline{G'_{0k}} = -\hbar c\,\overline{G_{0k}} + \frac{g_2}{\hbar c} i\overline{\psi}\beta\alpha^k\tau_{PN}\psi, \tag{27}$$

and

$$G_k \equiv \frac{1}{\hbar}\frac{\partial L}{\partial\left(\dfrac{\partial \overline{U_k}}{\partial t}\right)} = -G'_{0k} = -\hbar c\,G_{0k} + \frac{g_2}{\hbar c} i\overline{\psi}\beta\alpha^k\tau_{NP}\psi,$$

so that following Pauli and Weisskopf (1934) the commutation rules may be taken as

$$i[\overline{G_k}(x,t),\ U_l(x',t)] \equiv i\{\overline{G_k}\,U_l - U_l\overline{G_k}\} = \delta(x-x')\,\delta_{kl}$$

and

$$i[G_k(x,t),\ \overline{U_l}(x',t)] = \delta(x-x')\,\delta_{kl}, \tag{28}$$

with all other combinations commuting. The charge density given by (19) may be written

$$\sigma_0 = ie(\overline{U_k}G_k - U_k\overline{G_k}), \tag{29}$$

which in consequence of the commutation relations (28) has eigenvalues which are positive or negative integral multiples of e as has been shown by Pauli and Weisskopf.

From the equations (28) it may be deduced that

$$i[\overline{G_k}(x'),\ G_{lm}(x)] = \left\{\delta_{km}\left(\frac{\partial}{\partial x_l} - \frac{ie}{\hbar c}\phi_l\right) - \delta_{kl}\left(\frac{\partial}{\partial x_m} - \frac{ie}{\hbar c}\phi_m\right)\right\}\delta(x'-x), \tag{30a}$$

$$[G_k(x'),\ G_{lm}(x)] = 0, \tag{30b}$$

and the conjugate complex equations. The commutation rules for U_0 will be discussed below.

The commutation rules for ψ, $\overline{\psi}$ are as usual

$$\begin{aligned} \psi(x).\overline{\psi}(x')+\overline{\psi}(x').\psi(x) &= \delta(x-x').1 \\ \psi(x).\psi(x')+\psi(x').\psi(x) &= 0 \\ \overline{\psi(x)}.\overline{\psi(x')}+\overline{\psi(x')}.\overline{\psi(x)} &= 0, \end{aligned} \tag{31}$$

where 1 is a unit matrix with eight rows and columns corresponding to the eight rows of ψ.

Finally, the commutation rules for the electromagnetic field may be written in the form

$$[E_k(x'),\phi_l(x)] = 4\pi i c\hbar\,\delta_{kl}\delta(x-x'), \tag{32a}$$

from which follow, using (8),

$$[E_k(x'),H_{lm}(x)] = 4\pi i c\hbar\left\{\delta_{kl}\frac{\partial}{\partial x_m}-\delta_{km}\frac{\partial}{\partial x_l}\right\}\delta(x-x'), \tag{32b}$$

and
$$[\phi_k,H_{lm}] = 0.$$

The quantities describing either the electromagnetic field, or the proton-neutron, or the U-particles of course commute with those describing any other of the above fields.

The Hamiltonian for the whole system may then be written in the form

$$\mathscr{H} = \iiint dx_1dx_2dx_3(H^M+H^U+H^I+H^{\text{Max}}+H^0), \tag{33}$$

where, consistently with (26), H^M is the Hamiltonian for the heavy particles including their interaction with the electromagnetic field given by

$$H^M = \overline{\psi}\left[\alpha^k c\left(p_k-\frac{e}{c}\phi_k\tau_P\right)+\beta c^2(M_N\tau_N+M_P\tau_P)\right]\psi. \tag{34}$$

H^U is the Hamiltonian for the free U-particles together with their interaction with the electromagnetic field and a part of their interaction with the proton-neutron, being the term in curly brackets in (26), which by (27) becomes

$$\begin{aligned} H^U = {}&\left(\overline{G_k}+\frac{g_2}{e\hbar c}\psi^+\gamma^4\gamma^k\tau_{PN}\psi\right)\left(G_k+\frac{g_2}{e\hbar c}\psi^+\gamma^4\gamma^k\tau_{NP}\psi\right) \\ &+\tfrac{1}{2}\hbar^2c^2\overline{G_{kl}}G_{kl}+M_u^2c^4(\overline{U_k}U_k+\overline{U_0}U_0). \end{aligned} \tag{35}$$

I is the Hamiltonian for the interaction of the U-particles with the proton-neutron given by

$$I = g_1\psi^+\gamma^k(U_k\tau_{PN}+\overline{U_k}\tau_{NP})\psi+\tfrac{1}{2}g_2\psi^+\gamma^k\gamma^l(G_{kl}\tau_{PN}+\overline{G_{kl}}\tau_{NP}\psi). \tag{36}$$

$H^{\max}$ is the usual Hamiltonian for the electromagnetic field*

$$H^{\mathrm{Max}} = \frac{1}{8\pi}(\mathbf{E}^2 + \mathbf{H}^2). \tag{37}$$

Without going into the well-known complications which arise in the quantization of fields, I would only remark that H^0 is just the vanishing term which must be added to the Hamiltonian in order that the right equations of motion should result by an application of the rule

$$i\hbar\frac{\partial f}{dt} = [f, \mathscr{H}], \tag{38}$$

where f is any operator not explicitly involving the time. Here H^0 contains terms in G_k and U_0 similar to those in the electromagnetic field. We have

$$\begin{aligned} H^0 = \phi_0\{e\overline{\psi}\tau_P\psi + ie(\overline{U_k}G_k - U_k\overline{G_k}) - \tfrac{1}{4}\pi \operatorname{div}.\mathbf{E}\} \\ + \overline{U_0}\left\{\hbar c\left(\frac{\partial}{\partial x_k} - \frac{ie}{\hbar c}\phi_k\right)G_k - M_u^2c^4U_0 - g_1\overline{\psi}\tau_{NP}\psi\right\} + \text{conj}. \end{aligned} \tag{39}$$

The last terms also vanish on account of (15) and (27).

The commutation rules (28), (30), (31) and (32) suffice to derive the equations of motion of the whole system completely by an application of (38) to the various quantities concerned, treating U_0 and $\overline{U_0}$ as if they commuted with all quantities. The reason why U_0 and $\overline{U_0}$ behave *as if* they commuted with all quantities concerned will be discussed below.

Writing ψ, $\overline{\psi}$ in place of f we get just the equations (10) and (11) respectively.

Similarly, an application of (38) to ϕ_k and E_k leads just to the equations (8) and (16). The equation (17*b*) has as usual to be imposed as an initial condition at time $t = 0$.

Finally, applying the rule (38) to U_k,

$$i\hbar\dot{U}_k = i\left\{G_k + \frac{g_2}{e\hbar c}\psi^+\gamma^4\gamma^k\tau_{NP}\psi\right\} + e\phi_0 U_k - i\hbar c\left(\frac{\partial}{\partial x_k} - \frac{ie}{\hbar c}\phi_k\right)U_0,$$

which, remembering (27), are just three of the equations (14). Similarly, writing G_k for f in (38),

$$-i\hbar\dot{G}_k = -\hbar c\left(\frac{\partial}{\partial x_l} - \frac{ie}{\hbar c}\phi_l\right)G'_{lk} + M_u^2c^4U_k + g_1\psi^+\gamma^k\tau_{NP}\psi + ie\phi_0G_k,$$

---

* W. Heisenberg and W. Pauli (1929*a*, *b*). Cf. also W. Pauli (1933).

which are in fact just three of the equations (15) putting $\nu = k$. The fourth equation of the set (15a) got by putting $\nu = 4$, namely

$$\hbar c\left(\frac{\partial}{\partial x_k} - \frac{ie}{\hbar c}\phi_k\right)G_k - M_u^2 c^4 U_0 - g_1 \overline{\psi}\tau_{NP}\psi = 0 \tag{40}$$

and its conjugate equation in (15b) cannot be derived as equations of motion. This is obvious since they contain the operators G_k and U_0 but no time derivatives, so that they must be fulfilled at any given instant of time. They have to be imposed as initial conditions at $t = 0$, as, indeed, is also the case with equation (17b).

It is now possible to discuss the commutation rules for U_0 and $\overline{U}_0$. These occur in the Hamiltonian (33) only in the expression

$$M_u^2 c^4 \overline{U}_0 U_0 + \overline{U}_0\left\{\hbar c\left(\frac{\partial}{\partial x_k} - \frac{ie}{\hbar c}\phi_k\right)G_k - M_u^2 c^4 U_0 - g_1 \overline{\psi}\tau_{NP}\psi\right\} + \text{conj.} \tag{41}$$

The result, therefore, of some operator, say f, not commuting with U_0 or $\overline{U}_0$ is only to add to the expression for $[f, \mathscr{H}]$ terms containing the expression in curly brackets in (41) or its conjugate as factors on the right, which on account of the initial condition (40) vanish. This is the basis of what has been said before, namely that the commutation properties of U_0 and $\overline{U}_0$ are irrelevant in deriving the equations of motion, and these may be taken as commuting with all other quantities for this purpose.

It is clear, however, that U_0 and $\overline{U}_0$ cannot really commute with all quantities in $\mathscr{H}$, since this would mean by (38) that $\dot{U}_0$ and $\dot{\overline{U}}_0$ were zero. The commutation properties of U_0 and $\overline{U}_0$ can be derived from the fact that if the equations (40) and its conjugate are imposed at the time $t = 0$, they must be preserved for all time by the equations of motion. In other words, the expression on the left-hand side of (40) must commute with $\mathscr{H}$, and it must also commute with the corresponding conjugate expression. Denote by χ the expression on the left-hand side of (40) except the term $M_u^2 c^4 U_0$. Then

$$[\chi - M_u^2 c^4 U_0, \mathscr{H}] = 0, \tag{42}$$

with its conjugate equation. Indeed (42) demands that

$$\begin{aligned} i\hbar\frac{\partial}{\partial t}(M_u^2 c^4 U_0) &= [M_u^2 c^4 U_0, \mathscr{H}] = [\chi, \mathscr{H}] \\ &= g_1 \psi^+[ic^2(M_P - M_N)\tau_{NP} + (g_1 U_\mu \gamma^\mu + \tfrac{1}{2}g_2\gamma^\mu\gamma^\nu G_{\mu\nu})(\tau_P - \tau_N)]\psi \\ &\qquad - e\hbar c(\tfrac{1}{2}F_{kl}G'_{kl} + F_{0k}G_k) - M_u^2 c^4 i\hbar c\left(\frac{\partial}{\partial x_k} - \frac{ie}{\hbar c}\phi_k\right)U_k \\ &\qquad + eM_u^2 c^4 \phi_0 U_0. \end{aligned} \tag{43}$$

This equation is, of course, just what could be derived from the equations (15). Applying the operator $\left(\frac{\partial}{\partial x_\nu} - \frac{ie}{\hbar c}\phi_\nu\right)$ to the equations (15) and summing over ν,

$$M_u^2 c^4\left(\frac{\partial}{\partial x_\nu} - \frac{ie}{\hbar c}\phi_\nu\right) U_\nu = \frac{ie}{2} F_{\mu\nu} G'_{\mu\nu} - \left(\frac{\partial}{\partial x_\nu} - \frac{ie}{\hbar c}\phi_\nu\right)(g_1 \psi^+ \gamma^\nu \tau_{NP} \psi), \quad (43a)$$

which by using the equations (10) can be reduced to (43). Since the equations of motion do not depend on the commutation rules of U_0 and $\overline{U}_0$, the latter may clearly be chosen to satisfy (42), and its conjugate equation.

We therefore assume that the commutation rules of $M_u^2 c^4 U_0$ with all other quantities are exactly those of χ, so that (42) is satisfied identically. Indeed, this procedure is equivalent to treating (40) and its conjugate as the definition of U_0 and $\overline{U}_0$. The introduction of these quantities is, however, convenient for the general theory because it makes the relativistic invariance of the equations obvious.

Using (40) and its conjugate, H^U becomes

$$\begin{aligned} H^U = &\left(\overline{G}_k + \frac{g_2}{\epsilon\hbar c}\psi^+\gamma^4\gamma^k\tau_{PN}\psi\right)\left(G_k + \frac{g_2}{\epsilon\hbar c}\psi^+\gamma^4\gamma^k\tau_{NP}\psi\right) \\ &+ \tfrac{1}{2}\hbar^2 c^2 \overline{G_{kl}} G_{kl} + M_u^2 c^4 \overline{U}_k U_k \\ &+ \frac{\hbar^2 c^2}{M_u^2 c^4}\left\{\left(\frac{\partial}{\partial x_k} + \frac{ie}{\hbar c}\phi_k\right)\overline{G_k} - \frac{g_1}{\epsilon\hbar c}\psi^+\gamma^4\tau_{PN}\psi\right\} \\ &\times\left\{\left(\frac{\partial}{\partial x_k} - \frac{ie}{\hbar c}\phi_k\right)G_k - \frac{g_1}{\epsilon\hbar c}\psi^+\gamma^4\tau_{NP}\psi\right\}. \end{aligned} \quad (35a)$$

Now one may go over by a well-known procedure* to a state where there are only $1 \ldots s \ldots n$ proton-neutrons in the field. In this case in virtue of the commutation rules (31),

$$\overline{\psi}\gamma\tau\psi = \sum_{(s)} \gamma^{(s)}\tau^{(s)}\delta(x - X^{(s)}), \quad (44a)$$

where $\gamma^{(s)}, \tau^{(s)}$ and $X^{(s)}$ are the corresponding matrices and co-ordinates which refer to the sth particle. It is then obvious that each of those terms in (35a) which is a product of the type $(\overline{\psi}\gamma\tau\psi)(\overline{\psi}\gamma'\tau'\psi)$ gives rise to n infinities in the Hamiltonian (33) when (35a) is integrated over the whole of space, each infinity resulting according to (44) from cross terms in which the same suffix s appears twice. These terms represent just the infinite self energy of the proton-neutrons due to their interaction with the U-particles, and have to be neglected as must the electromagnetic self-energies. In particular, these terms are

---

* Cf. V. Fock (1932).

$$-\frac{g_1^2}{M_u^2c^4}(\psi^+\gamma^4\tau_{PN}\psi)(\psi^+\gamma^4\tau_{NP}\psi)-\frac{g_2^2}{\hbar^2c^2}(\psi^+\gamma^4\gamma^k\tau_{PN}\psi)(\psi^+\gamma^4\gamma^k\tau_{NP}\psi)$$

$$=\sum_{s\neq t}\tau_{PN}^{(s)}\overset{*}{\tau}{}_{NP}^{(t)}\left(\frac{g_1^2}{M_u^2c^4}+\frac{g_2^2}{\hbar^2c^2}\gamma^{k(s)}\gamma^{k(t)}\right)\delta(\mathbf{X}^{(s)}-\mathbf{X}^{(t)}). \qquad (44b)$$

$$+\text{self-energy terms.}$$

There is, in other words, an exchange interaction between different particles of the form of δ functions. The term in g_1^2 *exactly compensates* a δ function which appears in the second order interaction given in (67). These δ function interactions are however to some extent arbitrary. One could, for example, have started equally well from a Lagrange function in which the $\hbar cG_{\mu\nu}$ and $\hbar c\overline{G_{\mu\nu}}$ of (2) were replaced by $G'_{\mu\nu}$ and $\overline{G'_{\mu\nu}}$ of (14), at the same time omitting the g_2 term in (6). This is equivalent to adding

$$-\frac{1}{2}\frac{g_2^2}{\hbar^2c^2}(\psi^+\gamma^\mu\gamma^\nu\tau_{PN}\psi)(\psi^+\gamma^\mu\gamma^\nu\tau_{NP}\psi)$$

to the total Lagrangian L^U+L^I, or just minus this expression to the Hamiltonian (26). Such an addition leaves all the previous equations unchanged, and merely adds additional terms to the equations (10). Their effect is however to change the terms in g_2^2 in (44b) so that these now *exactly* cancel the corresponding δ function terms in (67). The reason why this is possible is essentially due to the fact that a δ-function is a relativistically invariant interaction and can therefore be added on at will. That, however, exactly the same spin dependence results as in (67) is not trivial.

For the purposes of calculation it is convenient to make a Fourier analysis of the U's and G's. It will be assumed as usual that all the functions are periodic in some very large volume V. Write

$$U_k=\frac{1}{\sqrt{V}}\sum_p u_{pk}e^{\frac{i}{\hbar}(\mathbf{p},\mathbf{x})} \qquad \overline{U_k}=\frac{1}{\sqrt{V}}\sum_p\overline{u_{pk}}e^{-\frac{i}{\hbar}(\mathbf{p},\mathbf{x})}$$

$$G_k=\frac{1}{\sqrt{V}}\sum_p g_{pk}e^{\frac{i}{\hbar}(\mathbf{p},\mathbf{x})} \qquad \overline{G_k}=\frac{1}{\sqrt{V}}\sum_p\overline{g_{pk}}e^{-\frac{i}{\hbar}(\mathbf{p},\mathbf{x})} \qquad (45)$$

It then follows from (28) that

$$\left.\begin{aligned}&i[\overline{g_{p'l}},u_{pk}]=i[g_{p'l},\overline{u_{pk}}]=\delta_{kl}\delta_{pp'},\\ \text{and}\quad&[\overline{g},\overline{u}]=[g,u]=[\overline{g},g]=[\overline{u},u]=0.\end{aligned}\right\} \qquad (46)$$

Introducing three real unit vectors ϵ_{1p}, ϵ_{2p}, ϵ_{3p} for each Fourier component, such that ϵ_{3p} is in the direction of p and the other two are perpendicular to it,

$$(\boldsymbol{\epsilon}_{3p}, \mathbf{p}) = |p|;\ (\boldsymbol{\epsilon}_{1p}, \boldsymbol{\epsilon}_{2p}) = (\boldsymbol{\epsilon}_{1p}, \boldsymbol{\epsilon}_{3p}) = 0$$
$$(\boldsymbol{\epsilon}_{1p}, \boldsymbol{\epsilon}_{1p}) = (\boldsymbol{\epsilon}_{2p}, \boldsymbol{\epsilon}_{2p}) = (\boldsymbol{\epsilon}_{3p}, \boldsymbol{\epsilon}_{3p}) = 1. \tag{47}$$

One may then resolve any vector v_k along the three directions $\epsilon_{1,2,3}$ thus

$$v_k = \sum_{r=1,2,3} \epsilon_{rpk}\, v_{rp},$$

so that

$$v_{rp} = \sum_k \epsilon_{rpk}\, v_k. \tag{48}$$

In particular, the u_{pk} and g_{pk} of (45) may be resolved into three components in the same way. The total Hamiltonian for the U-particles may then be written in the form

$$\mathscr{H}^U + \mathscr{I} \equiv \iiint dx_1\, dx_2\, dx_3 (H^U + I) = \mathscr{H}^U_0 + \mathscr{H}^U_1 + \mathscr{I}_0 + \mathscr{I}_1, \tag{49}$$

where $\mathscr{H}^U_0$ is the Hamiltonian for the free U-particles without any interaction with the proton-neutron or the electromagnetic field. Hence

$$\mathscr{H}^U_0 = \sum_p \left\{ \sum_{r=1,2} (\overline{g_{rp}} g_{rp} + E^2 \overline{u_{rp}} u_{rp}) + \frac{E^2}{M_u^2 c^4} \overline{g_{3p}} g_{3p} + M_u^2 c^4 \overline{u_{3p}} u_{3p} \right\}, \tag{50}$$

where

$$E = +c(p^2 + M_u^2 c^2)^{\frac{1}{2}}. \tag{51}$$

$\mathscr{H}^U_1$, $\mathscr{I}_0$, $\mathscr{I}_1$, will be given below. Following Pauli and Weisskopf and introducing quantities a, b and their conjugates defined by

$$g_{1p} = \sqrt{\frac{E}{2}} (a_{1p} + \overline{b_{1p}}), \quad u_{1p} = \frac{i}{\sqrt{(2E)}} (a_{2p} - \overline{b_{1p}}),$$
$$\overline{g_{1p}} = \sqrt{\frac{E}{2}} (\overline{a_{1p}} + b_{1p}), \quad \overline{u_{1p}} = -\frac{i}{\sqrt{(2E)}} (a_{1p} - \overline{b_{1p}}), \tag{52}$$

it follows from (46) that

$$[a_{1p'}, \overline{a_{1p}}] = [b_{1p'}, \overline{b_{1p}}] = \delta_{pp'}$$
$$[a_{1p'}, a_{1p}] = [b_{1p'}, b_{1p}] = [a_{1p'}, b_{1p}] = [a_{1p'}, \overline{b_{1p}}] = 0. \tag{53}$$

Similar equations hold for the a's and b's with the suffix 2, while for the longitudinal waves it is more convenient to introduce a_{3p}, b_{3p} by

$$g_{3p} = \frac{M_u c^2}{\sqrt{(2E)}} (a_{3p} + \overline{b_{3p}}); \quad u_{3p} = \frac{i}{M_u c^2} \sqrt{\frac{E}{2}} (a_{3p} - \overline{b_{3p}}),$$
$$\overline{g_{3p}} = \frac{M_u c^2}{\sqrt{(2E)}} (\overline{a_{3p}} + b_{3p}); \quad \overline{u_{3p}} = -\frac{i}{M_u c^2} \sqrt{\frac{E}{2}} (\overline{a_{3p}} - b_{3p}), \tag{54}$$

so that (50) reduces to*

$$\mathscr{H}_0^U = \sum_p E\{\overline{a_{rp}}a_{rp} + \overline{b_{rp}}b_{rp} + 3\}.$$

a_{3p} and b_{3p} satisfy the same commutation relations as (53), and all quantities with different suffices always commute.

After some easy calculation it is found that

$$\begin{aligned}\mathscr{H}_1^U = \sum_{p,p'} [&-ec\{\phi_k^{p-p}(p_k+p'_k)(\overline{u_{p'l}}u_{pl}) - (\phi_k^{p-p'}u_{pk})(p_l\overline{u_{p'l}}) \\ &-(\phi_k^{p-p'}\overline{u_{p'l}})(p'_l u_{pl})\} - \frac{e}{M_u^2c^3}\{(\phi_k^{p-p'}\overline{g_{p'k}})pg_{3p} + (\phi_k^{p-p'}g_{pk})p'\overline{q_{3p'}}\} \\ &+e^2\{(\phi_k\phi_k)^{p-p'}(\overline{u_{p'l}}u_{pl}) - (\phi_k\phi_l)^{p-p'}\overline{u_{p'l}}u_{pk}\} \\ &+\frac{e^2}{M_u^2c^4}\{(\phi_k\phi_l)^{p-p'}\overline{g_{p'k}}g_{pl}\}],\end{aligned} \tag{55}$$

where

$$\phi_k^{p-p'} \equiv \frac{1}{V}\iiint_V \mathbf{dx}\,\phi_k e^{\frac{i}{\hbar}(\mathbf{p}-\mathbf{p}',\mathbf{x})} \tag{56}$$

It is convenient for the purposes of calculation to introduce a new constant g'_2 defined by

$$g'_2 = g_2\frac{M_u c}{\hbar}, \tag{57}$$

so that g'_2 now has the same dimensions as g_1, namely $(\hbar c)^{\frac{1}{2}}$. The interaction of the U-particles with the proton-neutron is then given by

$$\begin{aligned}\mathscr{I}_0 = \sum_s\sum_{\mathbf{p}}\frac{1}{\sqrt{V}\sqrt{(2E)}}\Bigg\{\Bigg[&-ig_1\Big\{\sum_{r=1,2}\alpha_{rp}^{(s)}(\overline{a_{rp}}-b_{rp}) - \frac{p}{M_uc}(\overline{a_{3p}}+b_{3p}) \\ &+\alpha_{3p}^{(s)}\frac{E}{M_uc^2}(\overline{a_{3p}}-b_{3p})\Big\} - ig'_2\beta^{(s)}\Big\{\sum_{r=1,2}\alpha_{rp}^{(s)}\frac{E}{M_uc^2}(\overline{a_{rp}}+b_{rp}) \\ &+\sum_{r=1,2}\alpha_{3p}^{(s)}\alpha_{rp}^{(s)}\frac{p}{M_uc}(\overline{a_{rp}}-b_{rp}) + \alpha_{3p}^{(s)}(\overline{a_{3p}}+b_{3p})\Big\}\Bigg]\tau_{NP}^{(s)}e^{-\frac{i}{\hbar}(\mathbf{p},\mathbf{x}^{(s)})} \\ &+\Bigg[ig_1\Big\{\sum_{r=1,2}\alpha_{rp}^{(s)}(a_{rp}-\overline{b_{rp}}) - \frac{p}{M_uc}(a_{3p}+\overline{b_{3p}}) + \alpha_{3p}^{(s)}\frac{E}{M_uc^2}(a_{3p}-\overline{b_{3p}})\Big\} \\ &-ig'_2\beta^{(s)}\Big\{\sum_{r=1,2}\alpha_{rp}^{(s)}\frac{E}{M_uc^2}(a_{rp}+\overline{b_{rp}}) + \sum_{r=1,2}\alpha_{3p}^{(s)}\alpha_{rp}^{(s)}\frac{p}{M_uc}(a_{rp}-\overline{b_{rp}}) \\ &+\alpha_{3p}^{(s)}(a_{3p}+\overline{b_{3p}})\Big\}\Bigg]\tau_{PN}^{(s)}e^{\frac{i}{\hbar}(\mathbf{p},\mathbf{x}^{(s)})}\Bigg\},\end{aligned} \tag{58a}$$

---

* A summation from 1 to 3 is always implied over all repeated Latin suffices.

with

$$\mathscr{I}_1 = \sum_s \sum_p \frac{e}{\sqrt{V}} \left[\left\{ \frac{ig_1}{M_u^2 c^4} \phi_l(X^{(s)}) g_{pl} + \sum_{k \neq l} \frac{g_2'}{M_u c^2} (\beta \alpha^k \alpha^l)^{(s)} \phi_k(X^{(s)}) u_{pl} \right\} \tau_{PN}^{(s)} e^{\frac{i}{\hbar}(\mathbf{p},\, \mathbf{X}^{(s)})} + \text{conj.} \right]. \quad (58b)$$

The expression (58*b*) is remarkable in that it contains a product of the variables describing the *U*-particles, the proton-neutron, and the electromagnetic field.

It is important to notice that both (35*a*) and (36) contain interaction terms in which differentials of the *U*'s and *G*'s appear, and which consequently contain additional constants of the dimensions of a length. This would mean that on this theory there would also be Heisenberg showers in which several particles are created in one elementary process, but with this difference, that whereas on Heisenberg's original theory these showers would have consisted mostly of electrons, on the present theory they would consist mainly of *U*-particles, a few electrons, and some protons or neutrons. This would also seem to agree with what little experimental evidence there is on the subject. The existence of such processes can be seen at once from the fact that in the interaction (58*a*) there are terms which increase proportionally to *p* or *E*, and, indeed, such terms are connected with the longitudinal *U*-waves alone due to the vector interaction in (36), and to the transverse *U*-waves alone due to the tensor interaction in (36).

Finally the expression for the current (19) in terms of the *a*'s and *b*'s for the case of free *U*-particles with no interaction is

$$\sigma_k = e \frac{cp_k}{E} \{ \overline{a_{rp}} a_{rp} + \overline{b_{rp}} b_{rp} - \overline{a_{rp}}\, \overline{b_{rp}} - a_{rp} b_{rp} + 3 \}$$

$$+ \frac{ep}{M_u c} \{ \overline{a_{3p}} (\epsilon_{rpk} \overline{b_{rp}}) + (\epsilon_{rpk} \overline{a_{rp}}) \overline{b_{3p}} + b_{3p} (\epsilon_{rpk} a_{rp}) + a_{3p} (\epsilon_{rpk} b_{rp}) \} \quad (59)$$

The current therefore contains terms which give rise to a rapid oscillatory motion as in the Dirac and Pauli-Weisskopf theory. The last term in (59) expresses the fact that there is a current due to the *U*-particles having a spin, which is also in the nature of a rapid oscillatory motion.

3. THE INTERACTION OF NEUTRONS WITH PROTONS

We now proceed to calculate the interaction between a neutron and a proton which results from the interaction of the heavy particles with the field of the U-particles. The method here employed is the usual method of quantum mechanics, consisting in essence of a development in powers of the constants $(g_1^2/\hbar^3c^3)$ and $(g_2'^2/\hbar^3c^3)$ which correspond in our theory to the fine structure constant $e^2/\hbar c$ of electrodynamics. This interaction appears already in the second approximation, and leads to a Heisenberg plus a Majorana force between a proton and a neutron, as is found from experiment, to fit which we have then to assume that the above constants are of the order of a tenth. Now this would make our method of calculation very bad, were it not for the fact that the next contribution to the proton-neutron force comes from a process of the sixth order, being thus $(g_1^2/\hbar^3c^3)^2$ times smaller than the one we have calculated. In the extreme relativistic case, however, the effect of the higher order processes cannot be considered as small, since the collision cross-section, for example, increases with the energy, due to the fact pointed out in the previous sections, that our theory leads to showers of Heisenberg's type. The expression for the second order process in the extreme relativistic case is however independent of the actual magnitude of the constants g_1 and g_2, so that if they were sufficiently small, there would clearly be some *relativistic* region where the interaction would be given to a good approximation by the second order process alone, with neglect of the processes of higher order. Now this case, although it may not be of direct practical importance, is of interest inasmuch as it gives us the mathematical generalizations of the Heisenberg and Majorana exchange force for the relativistic case.

As is usual in problems of this sort, it is convenient to carry out the calculation in the system of co-ordinates in which the centre of gravity of the two particles is at rest. There are therefore initially two heavy particles, one in a proton state, and the other in a neutron state, moving with equal and opposite momenta p_0, and we wish to calculate the differential effective cross-section dQ for their being scattered through the angle θ, the proton becoming a neutron and vice versa. The scattering takes place through the following intermediate states.

$$P(\mathbf{p}_0)+N(-\mathbf{p}_0)\rightarrow\left\{\begin{matrix}N(\mathbf{p}_0')+U^+(\mathbf{p}_0-\mathbf{p}_0')+N(-\mathbf{p}_0)\\ P(\mathbf{p}_0)+U^-(\mathbf{p}_0'-\mathbf{p}_0)+P(-\mathbf{p}_0')\end{matrix}\right\}\rightarrow N(\mathbf{p}_0')+P(-\mathbf{p}_0').$$

The quantities in brackets denote the momenta of the proton, neutron and U-particles in the corresponding state.

Then, as usual,

$$dQ=\frac{1}{16\pi^2}\frac{E_0^2V^2}{(\hbar c)^4}\sum_m\left|\frac{(f\,|\,\mathscr{H}_0\,|\,m)\,(m\,|\,\mathscr{H}_0\,|\,i)}{E_i-E_m}\right|^2d\Omega. \tag{60}$$

Here $d\Omega$ denotes the element of solid angle into which the scattering takes place, $E_0 \equiv c(p_0^2+M^2c^2)^{\frac{1}{2}}$ is the initial energy of either of the heavy particles, and we take the masses of the proton and neutron to be equal, say M. The letters i, m and f denote the initial intermediate and final states of the whole system. The momenta of the U-particles in the intermediate states is $\pm p$ where

$$p = |\mathbf{p}_0-\mathbf{p}_0'| = 2p_0 \sin \tfrac{1}{2}\theta, \tag{61}$$

so that
$$E_i - E_m = E = c(p^2+M_u^2c^2)^{\frac{1}{2}}. \tag{62}$$

Using (58 a), one finds after some calculation that the expression within the modulus in (60) can be written in the form of a matrix element $(f|W|i)$ with

$$\begin{aligned} W = -\tau_{NP}^{(1)}\tau_{PN}^{(2)} \frac{e^{\frac{i}{\hbar}(\mathbf{p},\, \mathbf{x}_2-\mathbf{x}_1)}}{VE^2} \Big[& \{g_1^2 - g_2'^2 \beta^{(1)} \alpha_{3p}^{(1)} \alpha_{sp}^{(1)} \beta^{(2)} \alpha_{3p}^{(2)} \alpha_{sp}^{(2)}\} \\ & \times \left\{ \alpha_{sp}^{(1)} \alpha_{sp}^{(2)} + \frac{p^2}{M_u^2c^2} + \frac{E^2}{M_u^2c^4} \alpha_{3p}^{(1)} \alpha_{3p}^{(2)} \right\} \\ & + g_1 g_2' \frac{p}{M_u c} \{(1-\alpha_{sp}^{(1)}\alpha_{sp}^{(2)})(\beta^{(2)}\alpha_{3p}^{(2)} - \beta^{(1)}\alpha_{3p}^{(1)})\} \Big] \\ & - [\text{Same expression with indices (1) and (2) interchanged}] \end{aligned} \tag{63}$$

Here the suffix s denotes a summation over the values 1 and 2, i.e. over the two transverse U-particle states for the given momentum p. The last term in (63) results from the fact that in accordance with (31) the wave-function for the two heavy particles (1) and (2) has to be antisymmetrical in the initial and final states.

First consider the non-relativistic form of (63). In this case $|E_i - E_m| = E$ exactly, since in the system chosen the centre of gravity is at rest. But in the non-relativistic case, i.e. when the heavy particles are moving with small velocity, this is still approximately true in any other system, since the change in energy of the heavy particles is small compared with E. Hence the most general non-relativistic interaction between the two heavy particles is just derived by summing (63) over all intermediate momenta p. We now introduce the matrices σ defined by

$$i\sigma_k = \alpha_l \alpha_m, \tag{64}$$

the suffices being interchanged cyclically, so that the σ's correspond to the well-known Pauli spin matrices. In the non-relativistic approximation one may then leave out all the terms containing α's in (63) excepting those products giving σ, and put $\beta = -1$. (63) then reduces to

$$W = -(\tau^{(1)}_{NP}\tau^{(2)}_{PN} + \tau^{(1)}_{PN}\tau^{(2)}_{NP}) \sum_p \frac{e^{\frac{i}{\hbar}(\mathbf{p},\, \mathbf{X}_2 - \mathbf{X}_1)}}{V} \frac{p^2}{E^2 M_u^2 c^2} [g_1^2 + g_2'^2(\sigma^{(1)}_{1p}\sigma^{(2)}_{1p} + \sigma^{(1)}_{2p}\sigma^{(2)}_{2p})]$$

$$= -(\tau^{(1)}_{NP}\tau^{(2)}_{PN} + \tau^{(1)}_{PN}\tau^{(2)}_{NP}) \sum_p \left[g_1^2 + g_2'^2(\boldsymbol{\sigma}^{(1)}, \boldsymbol{\sigma}^{(2)}) + \frac{g_2'^2 \hbar^2}{p^2} (\boldsymbol{\sigma}^{(1)}\,\mathrm{grad})\,(\boldsymbol{\sigma}^{(2)}\,\mathrm{grad}) \right] \frac{1}{V} \frac{p^2}{E^2 M_u^2 c^2} e^{\frac{i}{\hbar}(\mathbf{p},\, \mathbf{X}_2 - \mathbf{X}_1)}. \tag{65}$$

Now
$$\sum_p e^{\frac{i}{\hbar}(\mathbf{p},\, \mathbf{X}_2 - \mathbf{X}_1)} \frac{c^2 p^2}{V E^2} = \iiint \frac{d\mathbf{p}}{p^3}\left(1 - \frac{M_u^2 c^4}{E^2}\right) e^{\frac{i}{\hbar}(\mathbf{p},\, \mathbf{X}_2 - \mathbf{X}_1)}$$
$$= \delta(\mathbf{X}_2 - \mathbf{X}_1) - \frac{M_u^2 c^4}{4\pi\hbar^2} \frac{e^{-\lambda|\mathbf{X}_2 - \mathbf{X}_1|}}{|\mathbf{X}_2 - \mathbf{X}_1|}, \tag{66}$$

with $\lambda \equiv M_u c/\hbar$. Therefore

$$W = (\tau^{(1)}_{NP}\tau^{(2)}_{PN} + \tau^{(1)}_{PN}\tau^{(2)}_{NP}\Bigg\{\left[\frac{g_1^2}{\hbar^2 c^2} + \frac{g_2'^2}{\hbar^2 c^3}(\boldsymbol{\sigma}^{(1)}, \boldsymbol{\sigma}^{(2)}) - \frac{g_2'^2}{M_u^2 c^4}(\boldsymbol{\sigma}^{(1)}, \mathrm{grad})\,(\boldsymbol{\sigma}^{(2)}, \mathrm{grad})\right] \frac{1}{4\pi} \frac{e^{-\lambda|\mathbf{X}_2 - \mathbf{X}_1|}}{|\mathbf{X}_2 - \mathbf{X}_1|}$$
$$- \left[\frac{g_1^2}{M_u^2 c^4} + \frac{g_2'^2}{M_u^2 c^4}(\boldsymbol{\sigma}^{(1)}, \boldsymbol{\sigma}^{(2)})\right] \delta(\mathbf{X}_2 - \mathbf{X}_1)\Bigg\}. \tag{67}$$

The *interaction is therefore just of the required form consisting of Heisenberg and Majorana forces of the right sign so as to allow one to make the triplet state of the deuteron the lowest stable state. We would emphasize the fact that since only the squares of g_1 and g_2' enter into this expression, the sign of the Majorana force is beyond our control, and it is to be looked upon as a strong argument in favour of this theory that it allows only that sign of the force which actually occurs in nature.* There are additional terms consisting of δ functions which have already been discussed above. Finally, there is a term which depends on the spins of the heavy particles in the direction of their mutual separation. The effect of this term on the binding energy of the deuteron and on the scattering of neutrons has not yet been investigated.

To evaluate the relativistic scattering of protons by neutrons, it is necessary to evaluate the matrix element (63), summing over the two possible states of spin of both heavy particles in their initial and final states. This summation can be carried out by the usual methods, and leads to a complicated expression for the cross-section (60) containing terms in

g_1^4, $g_1^3 g_2'$, $g_1^2 g_2'^2$, $g_1 g_2'^3$ and $g_2'^4$. Therefore only the expression to which this reduces when g_2' is put equal to zero is given, since this is then the relativistic generalization of a pure Heisenberg force. Thus,

$$dQ = \frac{1}{8\pi}\frac{g_1^4}{\hbar^4 c^4}\frac{1}{E_0^2 E^4}\left[c^2 p_0^4 + \frac{c^4 p_0^4}{4}(1-\cos\theta)^2 - \frac{2(E_0^2 - E^2)\,p_0^4}{M_u^2}\sin^2\theta \right.$$
$$\left. + \{2M^2c^4 + c^2p_0^2(1+\cos\theta)\}^2 \frac{p_0^4}{M_u^4 c^4}(1-\cos\theta)^2\right] d\cos\theta. \quad (68)$$

In the general extreme relativistic case

$$cp_0 \gg M, \quad (69)$$

one gets

$$dQ = \frac{1}{32\pi}\frac{(g_1^4 + 2g_2'^4)}{\hbar^4 c^4}\frac{p_0^2}{M_u^4 c^6}(1+\cos\theta)^2\, d\cos\theta. \quad (70)$$

Putting g_2 equal to zero this goes over into the corresponding form of (68) for extremely high energies. Now the spin dependence of a pure Majorana force is given by the operator $\frac{1}{2}\{1+(\sigma^{(1)},\sigma^{(2)})\}$ so that (67) would reduce to a pure Majorana force if we put $g_1 = g_2'$. Thus (70) shows that a pure Majorana force scatters in the extreme relativistic case $\frac{3}{4}$ times as much as a pure Heisenberg force of the same magnitude, but the angular dependence of the scattering is identical, as might have been expected, for then neither particle conserves its spin in the collision, whatever its interaction.

Thus the relativistic calculation justified the result anticipated in a previous note (Bhabha 1934) that there would be a proton-neutron chain on the passage of a fast heavy particle through matter, though the angular distribution given by (70) is not nearly as asymmetrical as the incomplete treatment had led one to expect. If ϵ be the fraction of the kinetic energy of the moving particle communicated to the other particle in the system in which it was at rest, then

$$\epsilon = \tfrac{1}{2}(1+\cos\theta), \quad (71)$$

and the distribution (70) is proportional to $\epsilon^2 d\epsilon$, so that a rapid degradation of the energy follows resulting in a type of cascade shower in heavy particles alone. The total cross-section increases as p_0^2, which is connected, as already mentioned, with the phenomenon of the Heisenberg showers.

4. SCATTERING OF *U*-PARTICLES BY PROTON-NEUTRONS

We now proceed to apply the theory to a process which is of interest in connexion with the collision of heavy electrons with nuclei, namely their scattering by heavy particles. This process is exactly analogous to the

Compton scattering of light by electrons. Calculating again in the system in which the centre of gravity is at rest, the process is represented schematically below.

$$U^+(\mathbf{p}_0)+N(-\mathbf{p}_0)+[P(0)]\rightarrow\left\{\begin{array}{l} P(0)+[P(0)] \\ \\ \\ U^+(\mathbf{p}_0)+N(-\mathbf{p}_0) \\ \quad +U^+(\mathbf{p}_0')+N(-\mathbf{p}_0') \end{array}\right\}\rightarrow U^+(\mathbf{p}')+N(-\mathbf{p}')+[P(0)]$$

$$U^-(\mathbf{p}_0)+P(-\mathbf{p}_0)+[N(0)]\rightarrow\left\{\begin{array}{l} N(0)+[N(0)] \\ \\ \\ U^-(\mathbf{p}_0)+P(-\mathbf{p}_0) \\ \quad +U^-(\mathbf{p}_0')+P(-\mathbf{p}_0') \end{array}\right\}\rightarrow U^-(\mathbf{p}_0')+P(-\mathbf{p}_0')+[N(0)].$$

Here $U^+(\mathbf{p}_0)$ denotes a positive U-particle with momentum $\mathbf{p}_0$. The square brackets are meant to indicate that the particle is in a state of negative energy corresponding to the momentum indicated.

The calculation is straightforward but complicated. We shall give only the result here. Summing or averaging over the two possible spin states of the heavy particles and over the three possible spin states of the U-particles, one gets for dq, the cross-section for the scattering of the U-particle through an angle θ in the system in which the centre of gravity is at rest, the expression

$$dq=\frac{1}{24\pi\hbar^4c^4}\frac{1}{E_i^2(E_i^2-M^2c^4)^2}[(E_i^2+M^2c^4)(A^2+B^2)+4E_iMc^2AB \\ +(E_i^2-M^2c^4)C^2\cos\theta]\,d\cos\theta, \quad (72)$$

where

$$\left.\begin{aligned} A&=g_1^2\left\{2(E+E_N)\frac{E^2}{M_u^2c^4}+(E_N-2E)\right\}+6g_1g_2'E\frac{M}{M_u} \\ &\qquad +g_2'^2\left\{4(E_N+E)\frac{p^2}{M_u^2c^2}+3E_N\right\}, \\ B&=-3g_1^2Mc^2-6g_1g_2'\frac{c^2p^2+E_NE}{M_uc^2}-3g_2'^2Mc^2, \\ C&=g_1^2\left(cp-2E_N\frac{Ep}{M_u^2c^3}-2\frac{p^3}{M_u^2c}\right)-6g_1g_2'cp\frac{M}{M_u} \\ &\qquad -g_2'^2\left(4E_N\frac{Ep}{M_u^2c^3}+cp+4\frac{p^3}{M_u^2c}\right). \end{aligned}\right\} \quad (73)$$

Here $|p|$ denotes the momentum of the U-particle or heavy particle, $E = c(p^2+M_u^2c^2)^{\frac{1}{2}}$ and $E_N = c(p^2+M^2c^2)^{\frac{1}{2}}$ their respective energies, and $E_i \equiv E+E_N$.

In the extreme relativistic case $p \gg Mc$ this reduces to

$$dq = \frac{1}{24\pi}\frac{(g_1^2+2g_2'^2)^2}{\hbar^4c^4}\frac{p^2}{M_u^4c^6}(1+\cos\theta)\,d\cos\theta, \tag{74}$$

and in the non-relativistic case $p \ll Mc$, M_uc, to

$$dq = \frac{3}{4\pi}\frac{(g_1+g_2')^4}{\hbar^4c^4}\frac{M^2}{(M+M_u)^2(M_u+2M)^2c^4}\,d\cos\theta. \tag{75}$$

It is interesting to note that whereas (67) does not depend on the sign of g_1 relative to g_2', a comparison of (75) with experiment would allow one to determine the relative sign. The cross section is however very small, being of the order 10^{-28} cm.^{2}. The application of these results to cosmic radiation and the calculation of some of the other processes mentioned in a previous note (Bhabha 1938 *b*) will be carried out in another paper.

SUMMARY

A theory is developed based on the idea that the proton and neutron are two states of the same particle, which can go over from one state to the other by the emission of a charged particle of mass intermediate between those of the proton and electron, as originally suggested by Yukawa. These "U-particles" are described by four wave-functions. Quantization of the theory leads as usual to positive and negative U-particles with a spin of one unit. The U-particles are identified with the heavy electrons of cosmic radiation. The theory leads uniquely to short range forces of the Heisenberg and Majorana type of such a sort as to allow one to make the ground state of the deuteron the triplet state, *the sign of the Majorana force being not at our choice*. The range of the forces is connected with the mass of the U-particles as before, and demands a mass of some two hundred times the electron mass. The relativistic generalizations of a pure Heisenberg and pure Majorana force are given. The relativistic scattering of U-particles by protons and neutrons is calculated. The theory also leads to showers of Heisenberg's type, but consisting mainly of heavy electrons, a few electrons, and some heavy particles.

REFERENCES

Bhabha 1934 *Nature, Lond.*, **134**, 934.
— 1938*a* *Proc. Roy. Soc.* A, **164**, 257–94.
— 1938*b* *Nature, Lond.*, **141**, 117.
Blackett and Wilson 1937 *Proc. Roy. Soc.* A, **160**, 304.
Breit, Condon and Present 1936 *Phys. Rev.* **50**, 825–45.
Fock 1932 *Z. Phys.* **75**, 622–47. (Correction.) *Z. Phys.* **76**, 852.
Fröhlich and Heitler 1938 *Nature, Lond.*, **141**, 37.
Heisenberg and Pauli 1929*a* *Z. Phys.* **56**, 1–61.
— — 1929*b* *Z. Phys.* **59**, 168–90.
Kemmer 1938 *Nature, Lond.*, **141**, 116.
Neddermeyer and Anderson 1937 *Phys. Rev.* **51**, 884.
Nishina, Takeuchi and Ichimiya 1937 *Phys. Rev.* **52**, 1198.
Pauli 1933 *Handbuch der Physik*, 2nd ed. **24**, part **1**.
Pauli and Weisskopf 1934 *Helv. Phys. Acta*, **7**, 709–31.
Proca 1936 *J. phys. Radium*, **7**, 347–53.
Street and Stevenson 1937*a* *Phys. Rev.* **51**, 884.
— — 1937*b* *Phys. Rev.* **51**, 1005.
Stueckelberg 1937 *Phys. Rev.*, **52**, 41.
Tuve, Heydenberg and Hafstad 1936 *Phys. Rev.* **50**, 806–25.
Wenzel 1937*a* *Z. Phys.* **104**, 34–47.
— 1937*b* *Z. Phys.* **105**, 738–46.
Yukawa 1935 *Proc. Phys.-Math. Soc. Japan*, **17**, 48.
Yukawa and Sakata 1937 *Proc. Phys.-Math. Soc. Japan*, **19**, 1084–93.

Showers produced by the penetrating cosmic radiation

By W. Heitler†

H. H. Wills Physical Laboratory, University of Bristol

(*Communicated by N. F. Mott, F.R.S.—Received 7 March* 1938)

1. Introduction

In two recent papers by Fröhlich, Heitler and Kemmer (1938) and by Kemmer (1938) it has been shown that the properties of the nuclear particles proton and neutron can qualitatively be understood on the assumption that a proton (neutron) is capable of emitting a *heavy positive* (*negative*) *electron* (denoted by Y^+, Y^-), transforming itself at the same time into a neutron (proton). The theory of the heavy electron—its wave equation and its interaction with the nuclear particles—was built up in close analogy to the theory of light and its interaction with an electron. We now apply this theory to the passage of a heavy electron through matter, and we shall find that it leads to a qualitative explanation of a number of cosmic-ray facts connected with the penetrating radiation.‡

In applying the theory to collisions of *fast* heavy electrons with nuclei, there is, however, a serious difficulty from the start: From the discussion of the nuclear properties it has become evident that the theory in its present form can only claim validity for relative energies between the heavy electron and the nuclear particles not very much greater than the rest energy μc^2 of the heavy electron, i.e. up to at most a few times 10^8 e-volts. For higher energies the theory leads to serious mathematical difficulties (diverging self-energy, diverging nuclear forces of higher order, etc.). For cosmic rays the interesting region is just the one for energies greater than 10^8 e-volts. It may be justifiable, in spite of these facts, to apply the theory to cosmic-ray heavy electrons, for two reasons: In the first place it is to be expected that the processes derived from the theory for energies of the order μc^2 will exist also at higher energies and will preserve a number of their qualitative features. In the second place the theoretical cross-section obtained for these processes at energies of the order 10^8 e-volts will at least be right in the order of magnitude. On the other hand, we must not attach any significance to the way in which the cross-sections are found to depend on energy.

---

† Research Fellow of the Academic Assistance Council.

‡ In this paper we consider only processes connected with *charged* heavy electrons but not those involving neutral particles of mass μ (Neutretto's). The existence of the latter is, although very probable, not yet proved by direct experiments.

Even for comparatively small energies of only μc^2 the theory leads to a number of processes giving rise to the creation of various types of showers which all seem to have been observed in connexion with the penetrating component of cosmic rays. Because of the breakdown of the theory at high energies, we shall compute the cross-sections for these processes only for energies ϵ small compared with Mc^2 (M = protonic mass), although there would be no difficulty in extending the same calculations for the case $\epsilon > Mc^2$. The nuclear particles can then all be considered to be at rest and their recoil energy can be neglected.

According to the papers mentioned above, the heavy electron is represented by a relativistic vectorial wave function ϕ_α ($\alpha = 0, 1, 2, 3$), satisfying in the absence of an external field the Klein-Gordon equation and

$$\sum_\alpha \partial\phi_\alpha/\partial x_\alpha = 0.$$

For a given Lorentz system ϕ_α can be replaced by a transverse space vector $\boldsymbol{\phi}$ given by the transverse part of ϕ_1, ϕ_2, ϕ_3 with $\operatorname{div}\boldsymbol{\phi} = 0$ and a longitudinal space vector $\boldsymbol{\psi}$. ψ is defined by

$$\lambda\psi_i = \frac{1}{c}\phi_i^{(l)} + \frac{\partial\phi_0}{\partial x_i},$$

where $\phi_i^{(l)}$ is the longitudinal part of ϕ_1, ϕ_2, ϕ_3 ($\operatorname{curl}\phi^{(l)} = 0$) and

$$\lambda = \frac{\mu c}{\hbar}.$$

The interaction with the nuclear particles was then assumed to be

$$\left.\begin{aligned} H'_{\mathrm{I}} &= \frac{g}{\lambda}\Pi[\operatorname{div}\boldsymbol{\psi}]_0 + \text{conjugate term}, \\ H'_{\mathrm{II}} &= \frac{f}{\lambda}\Pi[(\boldsymbol{\sigma}\operatorname{curl}\boldsymbol{\phi})]_0 + \text{conjugate term}, \end{aligned}\right\} \tag{1}$$

where Π is the operator transforming a neutron into a proton and is zero when applied to a proton. The expressions (1) have to be taken at the position of the heavy particle (0). g and f are two constants with the dimensions of a charge and have been determined from the nuclear forces

$$\frac{g^2}{\hbar c} \doteq \frac{f^2}{\hbar c} \doteq \frac{1}{8}. \tag{2}$$

$1/\lambda = \hbar/\mu c$ is the range of the nuclear forces. $\boldsymbol{\sigma}$ is the spin vector of the heavy particle. The matrix elements for the absorption by a proton of a free negative heavy electron represented by a plane wave $\mathbf{j}\exp[i(\mathbf{pr})/\hbar c]$ ($\mathbf{p}$ is the momentum multiplied by c, $\mathbf{j}$ the unit vector of polarization) are then given by

$$H'_{\mathrm{I}} = i\frac{g}{\lambda}p\sqrt{\frac{2\pi}{\epsilon}}, \quad H'_{\mathrm{II}} = i\frac{f}{\lambda}\sqrt{\frac{2\pi}{\epsilon}}(\boldsymbol{\sigma}[\mathbf{jp}]), \tag{3}$$

$$\epsilon^2 = p^2 + \mu^2 c^4.$$

Apart from a factor -1, the matrix elements for the emission of a positive heavy electron by a proton are the same.

Since we also wish to discuss processes in which light quanta are involved, we also need the interaction of a heavy electron with the radiation field. This interaction has been given by Kemmer (1938). If $\mathbf{A}$ represents the vector potential and if terms proportional to A^2 are neglected the interaction can be written in the following form making use of the two vectors $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ only

$$H' = \frac{ie}{4\pi\hbar c}\int d\tau\left\{\operatorname{div}\boldsymbol{\psi}^*\left(\mathbf{A}, \boldsymbol{\psi} + \frac{1}{\lambda c}\dot{\boldsymbol{\phi}}\right)\right.$$
$$\left. + \left(\operatorname{curl}\boldsymbol{\phi}^*. \left[\mathbf{A}, \boldsymbol{\phi} - \frac{1}{\lambda c}\dot{\boldsymbol{\psi}}\right]\right) - \text{conjugate terms}\right\}. \tag{4}$$

In this formula $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ are the wave functions for the longitudinal and transverse heavy electrons respectively. H' gives rise to various types of transitions involving light quanta such as the emission of a light quantum by a heavy electron, creation and annihilation of a pair of heavy electrons with emission or absorption of a light quantum etc. The matrix elements for these transitions can easily be obtained by expanding ψ and ϕ into plane waves

$$\boldsymbol{\phi} = \Sigma q\mathbf{j}\sqrt{(4\pi c^2)}\, e^{i(\mathbf{pr})/\hbar c},$$
$$\boldsymbol{\psi} = \Sigma Q\mathbf{j}\sqrt{(4\pi c^2)}\, e^{i(\mathbf{pr})/\hbar c}.$$

The amplitudes q, Q have matrix elements for transitions in which the number of particles in the state $\mathbf{p}$ decreases by one positive or increases by one negative particle, whereas the conjugate amplitudes q^*, Q^* correspond to an absorption of a negative particle or emission of a positive one in the state $\mathbf{p}$. The expressions for q, Q, q^*, Q^* are given in the paper by Fröhlich, Heitler and Kemmer (1938) equations (24), (27). We now give the matrix elements of H' for the transitions considered in the following sections:

(*a*) A longitudinal heavy electron with momentum $\mathbf{p}$ emits a light quantum $\mathbf{k}$ ($k = \hbar\nu$) having then a momentum $\mathbf{p}' = \mathbf{p} - \mathbf{k}$. After the emission the heavy electron may be a longitudinal or transverse one. We first write down the terms corresponding to a longitudinal particle before and after the emission

$$H' = \frac{e\hbar c}{2}\sqrt{\frac{2\pi}{\epsilon\epsilon' k}}\left[\frac{p}{p'}(\mathbf{p}'\mathbf{e}) + \frac{p'}{p}(\mathbf{p}\mathbf{e})\right]. \tag{5a}$$

e is the direction of polarization of the light quantum.

(*b*) A pair (a positive and negative longitudinal heavy electron) is annihilated emitting a light quantum **k**. If the positive particle has a momentum **p**$''$ and the negative one a momentum **p** the matrix element is

$$H'' = -\frac{e\hbar c}{2}\sqrt{\frac{2\pi}{\epsilon\epsilon'' k}}\left[\frac{p}{p''}(\mathbf{p}''\mathbf{e}) + \frac{p''}{p}(\mathbf{p}\mathbf{e})\right]. \tag{5b}$$

(*c*) If the heavy electron in the state **p** is longitudinal but in the states **p**$'$, **p**$''$ is transverse the corresponding matrix elements are

$$H' = \frac{ie}{2\lambda}\sqrt{\frac{2\pi}{\epsilon\epsilon' k}}\left[\epsilon' p(\mathbf{e}\mathbf{j}') - \frac{\epsilon}{p}([\mathbf{j}'\mathbf{p}'][\mathbf{e}\mathbf{p}])\right], \tag{5c}$$

$$H'' = \frac{ie}{2\lambda}\sqrt{\frac{2\pi}{\epsilon\epsilon'' k}}\left[\epsilon'' p(\mathbf{e}\mathbf{j}'') + \frac{\epsilon}{p}([\mathbf{j}''\mathbf{p}''][\mathbf{e}\mathbf{p}])\right]. \tag{5d}$$

j$'$, **j**$''$ are the unit vectors of polarization for the transverse electron.

For the following calculations we shall apply Born's approximation. This method is, however, in our case objectionable. Born's approximation represents an expansion with respect to g^2 and is therefore only applicable if processes of higher order in g^2 are less probable than those of lower order. As it will be seen in Section 3, this is not always the case. On the other hand, Born's approximation will give at least a rough orientation about the order of magnitude of the effects to be expected and we do not aim at more in this paper.

2. CASCADE SHOWERS

For cosmic-ray energies the protons and neutrons in a heavy nucleus can be considered as free except for those processes for which an energy transfer to other nuclear particles is required by the conservation laws. We consider therefore first processes arising from an impact of a heavy electron with a free neutron or proton at rest. The simplest process of this sort is the following:

$$Y^+ + N = P + \hbar\nu \quad \text{or} \quad Y^- + P = N + \hbar\nu. \tag{6}$$

The total energy of the heavy electron (kinetic + rest energy) is taken up by the light quantum. The cross-section for this process can easily be computed by the methods familiar from the radiation theory. The cross-section is

$$\Phi = \frac{2\pi}{\hbar v} |V|^2 \rho, \tag{7}$$

where v is the velocity of the incident heavy electron, V the matrix element of the interaction between the heavy electron, the proton and light quantum for the transition in question and ρ the density of energy levels in the final state. The process is of second order and takes place in two stages:

(1) The negative heavy electron with momentum $\mathbf{p}$, energy ϵ emits the light quantum $\hbar\nu$ and has then a momentum p' and energy ϵ'

$$\mathbf{p}' = \mathbf{p} - \mathbf{k}, \quad \epsilon'^2 = p'^2 + \mu^2 c^4 \tag{8a}$$

($\mathbf{k}$ = momentum of light quantum, $k = \hbar\nu$).

(2) The heavy electron with momentum $\mathbf{p}'$ is absorbed by the proton. Since the nuclear particle can be considered as at rest, there is no intermediate state in which the proton has emitted a light quantum.

The process can however also take place in the following way: (1*a*) The proton emits a positive heavy electron with momentum $\mathbf{p}''$. (2*a*) The two heavy electrons with momenta $\mathbf{p}'$, $\mathbf{p}''$ are annihilated emitting the light quantum $\mathbf{k}$. $\mathbf{p}''$ is then given by

$$\mathbf{p}'' + \mathbf{p} = \mathbf{k} \quad \text{or} \quad \mathbf{p}'' = -\mathbf{p}', \quad \epsilon'' = \epsilon'. \tag{8b}$$

Denoting the initial, the two intermediate, and the final states by A, I, II, F the matrix element is

$$V = \frac{H_{A\mathrm{I}} H_{\mathrm{I}F}}{E_A - E_{\mathrm{I}}} + \frac{H_{A\mathrm{II}} H_{\mathrm{II}F}}{E_A - E_{\mathrm{II}}}, \tag{9}$$

where E_A, E_{I}, E_{II} are the energies in the initial and intermediate states respectively. Taking into account that $k = \epsilon$ because of the conservation of energy, we have for the denominators of (9)

$$E_A - E_{\mathrm{I}} = \epsilon - (k + \epsilon') = -\epsilon',$$
$$E_A - E_{\mathrm{II}} = \epsilon - (\epsilon + \epsilon'') = -\epsilon'' = -\epsilon'.$$

We assume that the incident heavy electron is a longitudinal one.

For $H_{A\mathrm{I}}$ and $H_{\mathrm{II}F}$ the matrix elements (5) must be inserted and for $H_{\mathrm{I}F}$, $H_{A\mathrm{II}}$ the matrix elements (3). $H_{\mathrm{I}F}$ is identical with $H_{A\mathrm{II}}$ but has the opposite sign. We thus obtain for the contribution to V arising from intermediate states in which the particles are longitudinal according to (8*b*)

$$V' = \frac{2\pi ie\hbar c}{2}\frac{g}{\lambda}\frac{1}{\sqrt{(\epsilon k)}}\frac{1}{\epsilon'}\frac{p'}{\epsilon'}\left[\frac{p}{p'}(\mathbf{p}'\mathbf{e})+\frac{p'}{p}(\mathbf{pe})+\frac{p}{p''}(\mathbf{p}''\mathbf{e})+\frac{p''}{p}(\mathbf{pe})\right]$$
$$= \frac{2\pi ie\hbar cg}{\lambda p}\frac{p'^2}{\epsilon\epsilon'^2}(\mathbf{pe}).$$

The contribution arising from intermediate states where the particles with momenta $\mathbf{p}', \mathbf{p}''$ are transverse is obtained in a similar way

$$V'' = -\frac{2\pi e\hbar cf}{\lambda}\frac{1}{\epsilon\epsilon'^2\mu c^2}\sum_{j'}\left[\epsilon' p(\mathbf{ej}') - \frac{\epsilon}{p}([\mathbf{j}'\mathbf{p}'][\mathbf{ep}])\right](\boldsymbol{\sigma}[\mathbf{j}'\mathbf{p}']).$$

In this formula the sum has to be taken over the two directions of $\mathbf{j}' \perp \mathbf{p}'$. The two parts of V differ by a factor i and give therefore additive contributions to the cross-section ($\sim |V|^2$). V', V'' have to be inserted in equation (7). In our case ρ is

$$\rho = \frac{k^2 d\Omega}{(2\pi\hbar c)^3} = \frac{\epsilon^2 d\Omega}{(2\pi\hbar c)^3},$$

where $d\Omega$ is the element of the solid angle for the light quantum. Denoting the angle $\sphericalangle(\mathbf{p},\mathbf{k})$ by θ, taking into account equation (8*a*), and taking the sum over the directions of polarization of the light quantum

$$\Sigma\,|(\mathbf{pe})|^2 = p^2\sin^2\theta,$$

the contribution of V' to the differential cross-section is

$$d\Phi' = \frac{g^2}{\hbar c}\frac{e^2}{\hbar c}\frac{1}{\lambda^2}\frac{\epsilon}{p}\sin^2\theta\frac{(p^2+\epsilon^2-2p\epsilon\cos\theta)^2}{4\epsilon^2(\epsilon-p\cos\theta)^2}\,d\Omega.$$

This formula shows that for all energies the angular distribution is practically a $\sin^2\theta$ law. Thus the light quantum can be *emitted in all directions*. The same qualitative result holds also for the contribution $d\Phi''$ arising from V' although the angular distribution is here more complicated. A large angular spread was to be expected since in a process arising from a collision of a particle with energy ϵ and a particle at rest with mass M the angular spread is always of the order ϵ/Mc^2. In our case M is the proton mass and since ϵ was assumed to be small compared with Mc^2 the angular spread is large.

For the discussion of the total cross-section we confine ourselves to a qualitative consideration of the two cases $p < \mu c^2$ and $p > \mu c^2$. In the first case we put $\epsilon \fallingdotseq \mu$ and neglect p. In the second case we put $p \fallingdotseq \epsilon$. Integrating over all angles the contribution Φ' is:

$$\left.\begin{aligned} \Phi' &= \frac{2\pi}{3}\frac{g^2}{\hbar c}\frac{e^2}{\hbar c}\frac{1}{\lambda^2}\cdot\frac{\mu c^2}{p} \quad (p < \mu c^2), \\ \Phi' &= \frac{8\pi}{3}\frac{g^2}{\hbar c}\frac{e^2}{\hbar c}\frac{1}{\lambda^2} \quad (p > \mu c^2). \end{aligned}\right\} \tag{10a}$$

The contribition Φ'' can be obtained in a similar way. After some calculation it is found that

$$\left.\begin{aligned} \Phi'' &= \frac{2\pi}{3}\frac{f^2}{\hbar c}\frac{e^2}{\hbar c}\frac{1}{\lambda^2}\frac{\mu c^2}{p} \quad (p < \mu c^2), \\ \Phi'' &= \frac{44\pi}{15}\frac{f^2}{\hbar c}\frac{e^2}{\hbar c}\frac{1}{\lambda^2}\left(\frac{\epsilon}{\mu c^2}\right)^2 \quad (p > \mu c^2). \end{aligned}\right\} \tag{10b}$$

The total cross-section is

$$\Phi = \Phi' + \Phi''. \tag{10c}$$

For small values of p Φ increases as $1/p$ with decreasing energy. Thus for small energies ($< 10^8$ e-volts) the heavy electron is highly absorbable and the probability for transforming itself into a light quantum is large. For higher energies Φ' is constant but Φ'' increases as ϵ^2 with the energy. This increase does certainly not correspond to reality. As we have mentioned in the introduction our whole theory breaks down for energies $\epsilon \gg \mu c^2$ and no significance must be attached to the formulae in this region. But it might be expected that they still give the right order of magnitude for energies of the order μc^2. Taking the mean value of the two formulae for $p \gtrless \mu c^2$ and putting $f = g$, $p \fallingdotseq \mu c^2$

$$\phi \fallingdotseq 15\frac{g^2}{\hbar c}\frac{e^2}{\hbar c}\frac{1}{\lambda^2} = \frac{1}{50}\frac{1}{\lambda^2}. \tag{11}$$

(11) is the cross-section per proton and has to be multiplied by Z to give the cross-section of a nucleus. According to (11) a heavy electron of 10^8 e-volts would travel about 2 m. of lead before it is transformed into a light quantum. This is certainly a reasonable order of magnitude.

The light quantum emitted by the heavy electron will at once create an *ordinary cascade shower*. For energies of the order 10^8 e-volts this cascade shower will consist in lead of about 1–5 electrons.

It has been observed by several authors that the penetrating radiation produces a number of ordinary cascade showers. Auger† and his co-workers have found that at a depth 30–70 m. below sea-level there still exist ordinary showers which seem to come from all directions. This is in qualitative agreement with our theory.†

---

† I am very much indebted to Professor Auger for his kind communication of these results.

3. PENETRATING SHOWERS

The theory also gives rise to an ordinary scattering of a heavy electron by a proton or neutron:

$$Y^- + P = P + Y^{1-} \quad \text{or} \quad Y^- + N = N + Y^{1-}. \tag{12}$$

The cross-section is again given by formula (7). The two stages of the process are: (1) Y^- is absorbed by the proton which is transformed into a neutron; (2) The neutron emits a negative heavy electron Y^{1-}. (For the scattering of Y^- by a neutron the order of absorption and emission has to be reversed. The result is exactly the same for the scattering by a proton or neutron.) There are no other intermediate states. Thus

$$E_A - E_{\mathrm{I}} = -\epsilon,$$

where ϵ is the energy of the incident heavy electron. For the matrix elements $H_{A\mathrm{I}}$ and $H_{\mathrm{I}F}$ insert the expression (3), thus, for a longitudinal particle neglecting the energy transfer to the nuclear particle

$$V = \frac{2\pi g^2}{\epsilon^2\lambda^2} p^2,$$

and since $\rho = p\epsilon/(2\pi\hbar c)^3$ the differential cross-section is

$$d\Phi = \frac{2\pi\epsilon}{\hbar c\, p} \frac{(2\pi)^2 g^4}{\epsilon^4\lambda^4} p^4 \frac{p\epsilon\, d\Omega}{(2\pi\hbar c)^3} = \left(\frac{g^2}{\hbar c}\right)^2 \frac{1}{\lambda^2} \frac{p^4\, d\Omega}{\epsilon^2(\lambda\hbar c)^2}.$$

The expression is independent of the angle of scattering and thus leads to a *large angle scattering*.—We have so far considered the absorption and re-emission of a longitudinal particle. It can easily be seen that the re-emitted particle can also be a transverse one, the matrix element being given by the f-term of equation (3). Putting $f = g$ and taking the sum over all spin directions σ the square of the matrix element is exactly the same for longitudinal and transverse particles and the summation over the directions of polarization for the emitted particle results simply in a factor 3. Thus the total cross-section per nuclear particle becomes

---

† Another mode of production of cascade showers by a heavy electron has been discussed by Bhabha (1938*a*). The heavy electron produces a fast secondary electron in a direct head-on collision with an extra nuclear electron. The latter produces subsequently a cascade shower. Since, however, fast secondary electrons are very rare, the probability of production of a cascade shower of even medium size is extremely small, in any case very much smaller than that due to the process discussed in our paper.

$$\Phi_1 = \frac{12\pi}{\lambda^2}\left(\frac{g^2}{\hbar c}\right)^2 \frac{p^4}{\epsilon^2(\mu c^2)^2}. \tag{13}$$

Before discussing this formula the cross-section for the multiple processes arising from an impact of a heavy electron with a nuclear particle is computed. This theory has the property which Heisenberg (1936) pointed out to be a necessary condition for the existence of multiple processes: it contains a universal length, namely $1/\lambda = \hbar/\mu c$. Probably, therefore, that at sufficiently high energies the process

$$Y_0^+ + N = P + Y^+ + Y^- \tag{14}$$

has the same probability as the simple scattering (12).

The cross-section for (14) can easily be obtained. The process is a three-stage process: (1) Y_0^+ is absorbed by the neutron; (2) the proton emits Y^+; (3) the neutron emits Y^-. There is, however, another set of intermediate states: (1a) the neutron emits Y^-; (2a) the proton emits Y^+; (3a) the neutron absorbs Y_0^+. The matrix element is therefore

$$V = \frac{H_{A\mathrm{I}}H_{\mathrm{I\,II}}H_{\mathrm{II}F}}{(E_A-E_{\mathrm{I}})(E_A-E_{\mathrm{II}})} + \frac{H_{A\mathrm{I}a}H_{\mathrm{I}a\,\mathrm{II}a}H_{\mathrm{II}aF}}{(E_A-E_{\mathrm{I}a})(E_A-E_{\mathrm{II}a})} = \left(\frac{g}{\lambda}\right)^3 \frac{(2\pi)^{\frac{3}{2}}}{\sqrt{(\epsilon\epsilon_+\epsilon_-)}}\,\frac{2pp_+p_-}{\epsilon\epsilon_-},$$

where ϵ_+, ϵ_- are the energies of the emitted particles, ϵ the energy of the incident particle and $\epsilon_+ + \epsilon_- = \epsilon$. The density function ρ is

$$\rho = \frac{p_+\epsilon_+p_-\epsilon_-\,d\epsilon_+\,d\Omega_+\,d\Omega_-}{(2\pi\hbar c)^6}.$$

Integrating over all angles and multiplying by a factor 9 to allow for all different directions of polarizations for the emitted particles, the total cross-section for the emission of a pair with an energy of the positive particle in the energy interval $d\epsilon_+$ becomes

$$\Phi = \frac{9\cdot 16}{\lambda^2}\left(\frac{g^2}{\hbar c}\right)^3 \frac{pp_+^3p_-^3}{\epsilon^2\epsilon_-^2(\mu c^2)^4}\,d\epsilon_+. \tag{15}$$

If all energies ϵ, ϵ_+, ϵ_- are comparable or larger than μc^2 we may approximately put $p_- = \epsilon_-$, $p_+ = \epsilon_+$. Integrating over ϵ_+ from 0 to ϵ we obtain for the total cross-section for the emission of two particles

$$\Phi_2 = \frac{36}{5}\frac{1}{\lambda^2}\left(\frac{g^2}{\hbar c}\right)^3\left(\frac{\epsilon}{\mu c^2}\right)^4. \tag{16}$$

Comparing this formula with Φ_1 equation (13) the ratio is

$$\frac{\Phi_2}{\Phi_1} = \frac{3}{5\pi}\frac{g^2}{\hbar c}\left(\frac{\epsilon}{\mu c^2}\right)^2, \tag{17}$$

and this ratio is of the order of magnitude one for $\epsilon \doteqdot 5\mu c^2$, the exact figure has, of course, no real significance. Thus at energies of the order μc^2 multiple processes take place with appreciable probability, and this will be the case also for processes of still higher order in which three or more particles are emitted. The number of particles is limited only by the energy of the primary particle. These processes will appear as *showers consisting of penetrating particles and produced by a penetrating particle.* Showers of this type have in fact been observed. They are responsible for the *second maximum of the Rossi curve.* The existence of this second maximum (which occurs at about 20 cm. of lead) shows directly that the penetrating particles are capable of producing secondary particles with a penetrating power of about 20 cm. of lead.

The *angular spread* of the penetrating showers is expected to be *large.* This does not seem, however, to be in agreement with experiments by Bothe and Schmeiser (1937) who concluded that the showers responsible for the second maximum of the Rossi curve have a very small angular spread. It is not possible to say at present whether this is the case for all penetrating showers. If their result were established it would lead to the conclusion that these showers are produced in collisions of a heavy electron with very *light* particles (i.e. electrons, the angular spread is then mc^2/ϵ). It might be that a heavy electron has also an interaction with an electron similar to that assumed in our theory between a heavy electron and a nuclear particle. This has in fact been suggested by Bhabha (1938*b*). A heavy electron could then produce a penetrating shower also in a collision with an electron. To settle this question definitely more extended experiments on the line of Bothe and Schmeiser are required.

It is not possible to draw any quantitative conclusions from our theory about the magnitude of the cross-section for the processes (12), (14), etc.† The theory breaks down for energies $> \mu c^2$. Indeed, according to (13) the cross-section for scattering would increase with ϵ^2 for higher energies and the "penetrating particles" would not be penetrating at all. On the contrary, it seems probable that Φ *decreases* with increasing energy.‡ Since, however, multiple processes occur even at energies of the order μc^2, this qualitative result—the existence of multiple processes—will probably prove to be correct. It is also to be expected that the order of magnitude of the cross-section is right for an energy $\epsilon = 2\mu c^2$ (kinetic energy $= \mu c^2$). From (13) is obtained a cross-section for an anomalous scattering $12\pi(g^2/\hbar c)^2\, 2/\lambda^2 \fallingdotseq 2/\lambda^2$ per proton or neutron leading to an average range of about 2–5 cm. of lead. Thus, only for energies $\epsilon > \mu c^2$ the particles will have great penetrating power. They become absorbable as soon as they reach an energy of the order 10^8 e-volts. A similar behaviour was found in Section II. It might be that this is one of the reasons why only very few "penetrating" particles are observed at lower energies.†

---

† No application can for instance be made of our theory to the energy loss curve measured by Blackett and Wilson (1937).

‡ The increase of the cross-section with ϵ^2 is due to the fact that the matrix element for the emission or absorption of a heavy electron is proportional to $p/\sqrt{\epsilon}$. This was also the reason for the diverging effects in the theory of the nuclear pro-

The cross-section for multiple processes would, at the same energy, be of the order $1/2\lambda^2$ per nuclear particle leading to the production of a penetrating shower after a range of 10–20 cm. of lead, which is in qualitative agreement with the facts.

4. SHOWERS OF HEAVILY IONIZING PARTICLES

So far we have considered processes in which the nuclear particles can be considered as free. The fact that the neutrons and protons in a heavy nucleus are bound leads to the following type of process: A heavy electron Y^+ with energy ϵ is absorbed by a bound neutron. The neutron gives all the energy to the nucleus which then becomes highly excited. According to Bohr's theory the energy ϵ is restored as "heat energy" in the nucleus. In course of time the nucleus evaporates, emitting subsequently a number of heavy particles, protons, neutrons and possibly also light quanta. (The latter would at once create pairs.) The particles can be emitted in all directions. This process will appear as a shower with angular spread 4π consisting of *heavy particles*. Showers of this type have been observed by several authors (cf. for instance, the beautiful photograph of Anderson and Neddermeyer (1936), fig. 10).

A quantitative treatment of this process is difficult because it depends on the details of the nuclear structure. It can be estimated that the number

---

perties. In order to obtain, for instance, a finite magnetic moment, it is at least necessary that, for energies $> \mu c^2$, the matrix element behaves like $1/\sqrt{\epsilon}$ instead of $\epsilon/\sqrt{\epsilon}$. Such a modification would at once lead to a $1/\epsilon^2$ law for the scattering cross-section and also for the multiple processes of higher order.

† According to Blackett (1938) the main reason for the absorption of heavy electrons is, however, their "transformation" into ordinary electrons. This would be very difficult to understand on any theory. In this paper we shall not discuss this question since it meets in any case with great difficulties.

of heavy particles emitted by a nucleus with an excitation energy of a few 10^8 e-volts ranges between two and ten (cf. Weisskopf 1937).

About the cross-section for this process only a few qualitative statements can be made: For very small energies—of the order of the binding energy W of the nuclear particle (i.e. a few 10^6 e-volts) the cross-section is found to be of the order $1/\lambda^2$ per nuclear particle. For higher energies the cross-section will decrease rapidly, with a high power of W/ϵ because the nuclear particles are then practically free. The rate of decrease depends very much on the model used for the nucleus. For cosmic-ray energies the cross-section is certainly very much smaller than $1/\lambda^2$, of the order $1/100\lambda^2$–$1/1000\lambda^2$. This is also in qualitative agreement with the observations.

Some of the observed showers containing heavy tracks with large angular spread may be due to multiple processes of the type discussed in the preceding section. The heavy tracks would then be tracks of comparatively slow heavy electrons. It is of course also quite possible that a combination of the processes discussed in Section 3 and a nuclear evaporation takes place. The heavy electron is captured by a nuclear particle and re-emitted with less energy (or a few heavy electrons are re-emitted). The nucleus remains excited and evaporates as described above. Then a shower is produced containing both proton and heavy electron tracks, as seems to have been observed by Anderson and Neddermeyer (1936).

5. ARE THE PENETRATING PARTICLES OF SECONDARY ORIGIN?

By the inverse process of (6)

$$h\nu + P = N + Y^+, \tag{18}$$

penetrating particles can be *created by light quanta*. For energies of the order μc^2 the cross-section for the inverse process is the same and given by equation (11).† The question arises now whether the penetrating particles observed at sea-level are perhaps all secondaries created by light quanta in the high atmosphere. The following consideration shows that the order of magnitude of the effect is quite sufficient to account for the number of penetrating particles observed at sea-level.

Light quanta are mainly absorbed by the creation of electron pairs. For energies of the order 10^8 e-volts the cross-section for pair creation is $10Z^2r_0^2/137$, where r_0 is the electronic radius. Comparing this with the cross-section (11) for the production of a heavy electron we find for air (putting $1/\lambda = r_0$)

---

† This cross-section is very much larger than that for the ordinary creation of a heavy electron pair by a light quantum in the presence of a nucleus.

$$\Phi_{\text{heavy electron}}/\Phi_{\text{pair}} \doteqdot 1/5Z \sim 1/40.$$

Thus about 1 in 40 light quanta will be absorbed by production of a heavy electron rather than by pair creation.

According to the measurements by Pfotzer (1936) the number of electrons in the high atmosphere (at 10 cm. Hg) is, in arbitrary units, 250, while the number of particles at sea-level is 7, most of which are penetrating particles. Most of the electrons in the high atmosphere have energies between 10^8 and 5×10^8 e-volts (cf. Heitler 1937, Nordheim 1937). The number of light quanta in a shower is always about twice as large as the number of electrons of the same energy; thus we have at least 500 light quanta in the high atmosphere. They will have created, therefore, at least 10 penetrating particles before they are absorbed. This figure is probably still too small for the following reason: it is seen in § 3 that at energies $> \mu c^2$ multiple processes occur. Presumably, therefore—and this can easily be verified by direct calculations—processes also will occur in which two or more heavy electrons are created by one light quantum

$$\hbar\nu + P = P + Y^+ + Y^-. \tag{19}$$

The cross-section for (19) will be not very much smaller than that for (18). The number of particles created is limited only by the energy of the light quantum. By the existence of these multiple processes the mean number of heavy electrons created by one light quantum will be increased perhaps by a factor 2. Thus we find that, in the above units, at least 20 heavy electrons are created in the high atmosphere. The measured absorption coefficient of the penetrating component is about 0·02/cm. Hg at sea-level. The number of penetrating particles to be expected at sea-level is therefore at least of the order

$$20 \exp(-0{\cdot}02 \times 65) \doteqdot 5,$$

as it is in fact observed. It is therefore concluded that it is quite possible to consider all *penetrating particles as secondaries produced by light quanta and electrons in the high atmosphere.* The same conclusion has been reached by Heitler (1938) from the consideration of the latitude effect of the penetrating radiation. We may mention that the cross-section for the production of heavy electrons is still small compared with that of ordinary pair creation. The absorption law of light quanta will therefore not be influenced by the creation of heavy electrons.

For a more quantitative treatment of this question it would be necessary to know the cross-section for the process (18) also at energies $> \mu c^2$, but this cannot be derived from the theory in its present stage. By the same reason it is not possible to predict the latitude effect of the penetrating radiation. The question could, however, easily be settled experimentally. If our considerations are right, the intensity of the penetrating component should have a maximum at a height slightly less than the maximum of the Regener curve, i.e. at about 20 cm. Hg.

SUMMARY

The theory of the heavy electron developed by Fröhlich, Heitler and Kemmer, has been applied to the passage of fast heavy electrons through matter. It is shown that various types of showers are produced (Y = heavy electron, P = proton, N = neutron).

(1) $Y^+ + N = P + h\nu$. The light quantum produces an ordinary cascade shower.

(2) Multiple processes of the type $Y^+ + N = P + Y^+ + Y^-$ and of higher order are possible. They lead to penetrating showers produced by penetrating particles.

(3) In a heavy nucleus: $Y^+ + N = P$. The energy of Y^+ is restored in the nucleus which subsequently evaporates emitting a few protons and neutrons (proton shower) accompanied possibly by electrons and also heavy electrons.

The inverse process to (1) leads to the creation of heavy electrons by light quanta. It is shown that the order of magnitude of the cross-section is sufficient to explain all heavy electrons at sea-level as secondaries produced in the high atmosphere.

Owing to the breakdown of the theory for energies $\gg \mu c^2$ (10^8 e-volts) all these processes can only be discussed qualitatively and their cross-section can only be calculated approximately for energies of the order 10^8 e-volts or less. For this energy the cross-sections have a reasonable order of magnitude.

REFERENCES

Anderson and Neddermeyer 1936 *Phys. Rev.* **50**, 263.
Bhabha 1938*a* *Proc. Roy. Soc.* A, **164**, 257.
Bhabha 1938*b* *Nature, Lond.*, **141**, 117.
Blackett and Wilson 1937 *Proc. Roy. Soc.* A, **160**, 304.
Blackett 1938 *Proc. Roy. Soc.* A (in the Press).
Bothe and Schmeiser 1937 *Naturwissenschaften*, **25**, 669, 833.

Fröhlich, Heitler and Kemmer 1938 *Proc. Roy. Soc.* A, **166**, 154.
Heisenberg 1936 *Z. Phys.* **101**, 533.
Heitler 1937 *Proc. Roy. Soc.* A, **161**, 261.
— 1938 *Report of the Moscow Conference.*
Kemmer 1938 *Proc. Roy. Soc.* A, **166**, 127.
Nordheim 1937 *Phys. Rev.* **51**, 1110.
Pfotzer 1936 *Z. Phys.* **102**, 23, 41.
Weisskopf 1937 *Phys. Rev.* **52**, 295.

The formation of helium molecules

By F. L. Arnot, Ph.D., *Lecturer in Natural Philosophy*
and Marjorie B. M'Ewen, B.Sc., *Demonstrator in Physics,*
The University, St Andrews

(*Communicated by H. S. Allen, F.R.S.—Received* 11 *March* 1938)

INTRODUCTION

The occurrence of band spectra in the chemically monatomic gases leads to the belief that molecules are formed in the gas by the electrons or photons producing the spectra. Arnot and Milligan (1936) and Arnot and M'Ewen (1938) have shown that diatomic mercury molecules are formed by the attachment of excited atoms in P-states to each other and to normal atoms. Although it has been inferred from the band spectra found in helium that molecules are formed by some such attachment process, these have not previously been studied electrically, nor have the details of the attachment process been investigated.* No definite evidence has been put forward to show that band spectra occur in the other rare gases.

This paper contains the results of an investigation made by the balanced space-charge method of the ionization produced in argon, neon and helium when these gases are subjected to bombardment by electrons whose energy was varied from a few volts below the first excitation potential to a few volts above the atomic ionization potential of the gas. Only atomic ionization has been found to occur in argon and neon, but molecular ions are found to be produced in helium when the energy of the electrons exceeds the resonance potential of the gas.

---

* In the early experiments on critical potentials ions were found to be produced in helium when the bombarding electrons had energies less than that corresponding to the atomic ionization potential. These effects were generally attributed to impurities in the helium used, but apparently never to the formation of helium molecules.

APPARATUS

The apparatus used in this work and the experimental procedure adopted have been fully described in a previous paper on the formation of mercury molecules (Arnot and M'Ewen 1938). In the present work we used the second form of apparatus shown in fig. 1*b*, as this was found in the work on mercury to be equally satisfactory and rather more sensitive than the form shown in fig. 1*a*. The wiring diagram is shown in fig. 1*a*. All metal parts, including gauzes, filament leads and connecting wires, were made of nickel, spot-welded where necessary. The cylinders C_1 and C_2 were of identical construction except for the gauze window in C_1. The caps on the ends of these cylinders were pressed out of one piece of nickel to give a perfect fit, and then were spot-welded to the cylinders. C_2 was supported from C_1 by quartz rods which insulated it from the rest of the apparatus. The filaments F_2 and F_3, which were connected in parallel, were co-axial with the cylinders C_1 and C_2, and their leads were insulated from the cylinders by short quartz tubes passing through the end-caps.

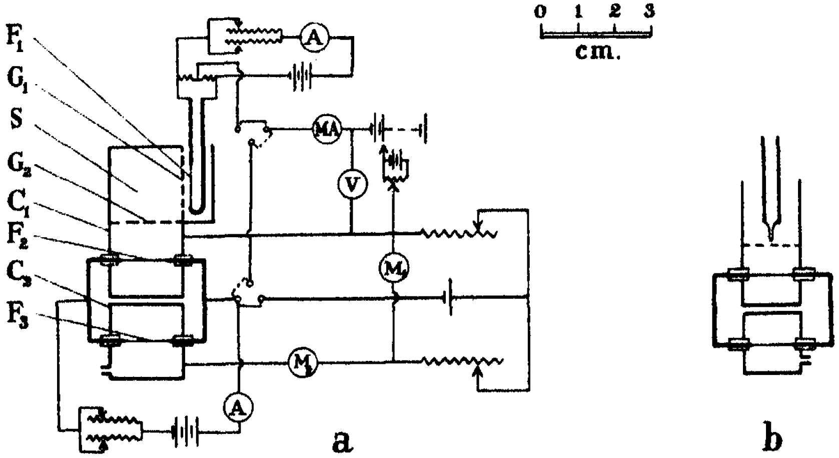


FIG. 1. Apparatus and wiring diagram.

The apparatus was contained in a large pyrex tube fitted with a 5 cm. ground-glass joint, so that the whole apparatus could be withdrawn for adjustment and filament renewal. The joint was water-cooled and a low vapour pressure Apiezon grease was used on this and all other joints and taps in the vacuum system. The tube was evacuated through a mercury cut-off and a liquid-air trap. The gases used were stored in a glass bulb of about 1 l. capacity, and were admitted to the apparatus through a fine glass capillary leak and a liquid-air trap. The pressure of gas in the apparatus, which was read on a McLeod gauge, could be altered by varying the pressure in the bulb and by adjusting the mercury cut-off, of which the inner tube was drilled with a number of holes of varying diameter arranged in the form of a spiral round the tube. The rare gases were supplied by the British Oxygen Co. in glass containers as spectroscopically pure.

EXPERIMENTAL PROCEDURE

After the usual baking-out, flashing and running-in of all filaments the current through the filaments F_2 and F_3 was set to give a saturated emission from each filament of about 0·1 mA. The potential between F_2 and F_3 and their respective cylinders was then reduced to 2 V resulting in the emissions from F_2 and F_3 being strongly space-charge limited to a value of about 0·01 mA. The emissions from F_2 and F_3 were then balanced by the bridge arrangement shown in fig. 1 which incorporated two 10,000 ohm resistance boxes and the galvanometer M_1. In this way any fluctuations in the potentials of the cells supplying the heating current to F_2 and F_3 were eliminated. The emission from F_3 alone was read by the galvanometer M_2 so that its constancy during a run could be checked.

Electrons from the filament F_1 are accelerated up to the gauze G_1 by a potential V_0 applied to the centre of the filament as shown in fig. 1, V_0 being read on a Weston standard voltmeter. When V_0 is increased to such a value that ions are formed in the cylinder C_1 their space-charge partially neutralizes the electron space-charge around F_2. The consequent increase in the emission from F_2 is read on the galvanometer M_1. This method is known to be a very sensitive detector of positive ions, since it has been shown that a single positive ion neutralizes the space-charge of as many as 10^4 to 10^6 electrons, for the ions revolve in spiral orbits around the filament, making many loops before finally hitting the fine wire.

RESULTS

The results obtained at various pressures in argon, neon and helium are shown respectively in figs. 2, 3 and 4. These curves represent the change

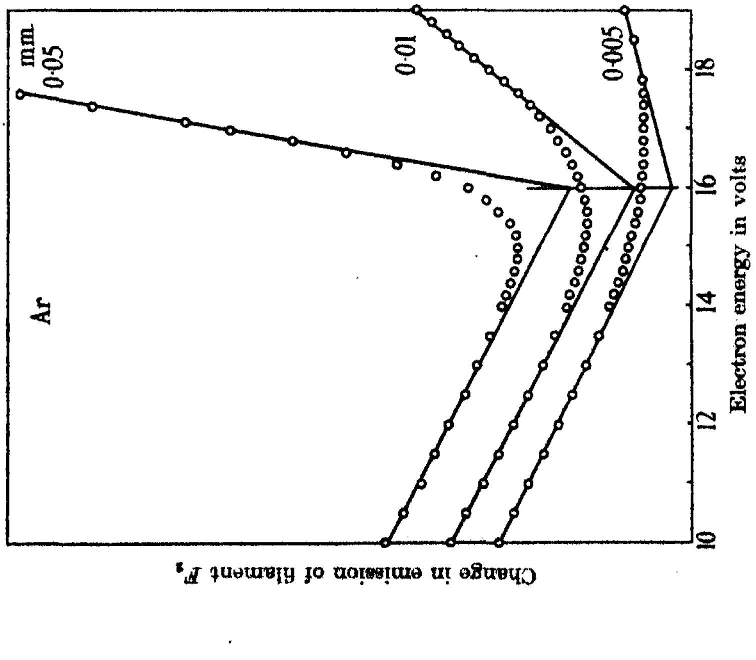


FIG. 2

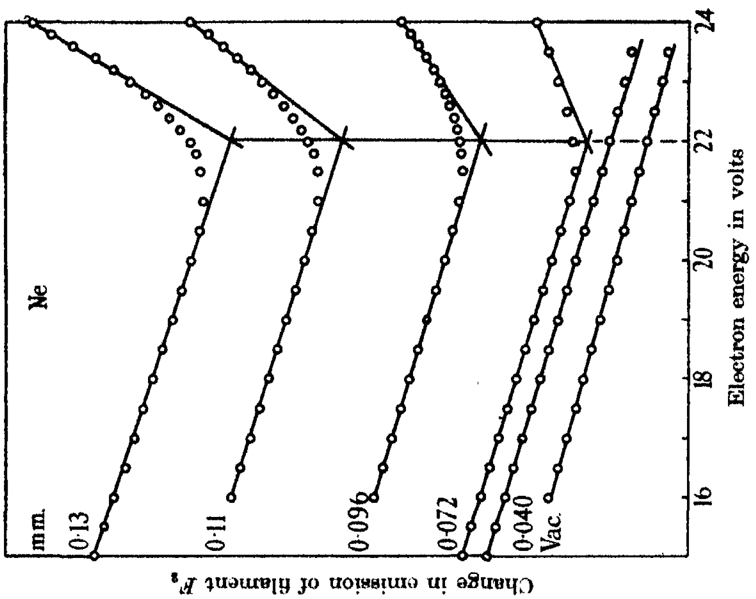


FIG. 3

FIG. 2. Showing onset of atomic ionization in argon. The spectroscopic value of the atomic ionization potential is 15·7 V and therefore the correction to the voltage scale for contact potentials is − 0·3 V.

FIG. 3. Showing onset of atomic ionization in neon. The spectroscopic value of the atomic ionization potential is 21·5 V and therefore the correction to the voltage scale for contact potentials is − 0·5 V.

in the space-charge limited current from the filament F_2, measured by the galvanometer M_1, as the energy V_0 of the electrons from F_1 is increased. The curves in figs. 2 and 3 have all been displaced in a vertical direction in order to separate clearly the points. The lowest curves in figs. 3 and 4 represent runs taken in vacuum. For these runs the McLeod gauge registered a "sticking vacuum" which is equivalent to a pressure of less than 10^{-6} mm. of Hg.

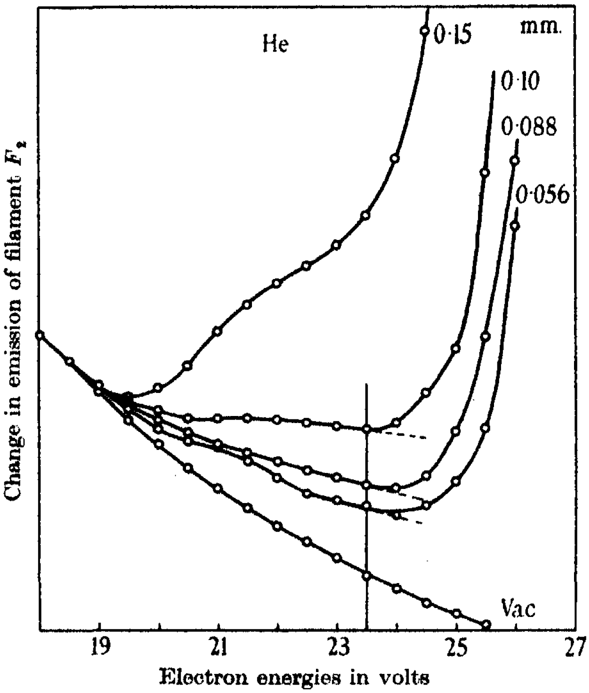


Fig. 4. Showing onset of molecular and atomic ionization in helium.

The linear fall in the curves as V_0 is increased up to the value at which ions begin to be formed has been investigated and discussed in our previous paper on the formation of mercury molecules (1938). The fall in the curves was there shown to be due to the fact that the emission from F_2 is not space-charge limited over its entire length. The temperature of the filament is highest in the centre, and the emission drawn from this portion is limited by space-charge; but the emission from the ends of the filament, which are cooled by the leads, is temperature-limited. Electrons from F_1 can have no effect on the space-charge limited emission from the centre of F_2, but they decrease the temperature-limited emission from the ends of F_2 by virtue of their space charge. Now as V_0 is increased the number of electrons penetrating the gauze G_1 and so affecting the emission from the ends of F_2 is increased, thereby causing the emission from F_2 to decrease as V_0 is increased.

It will be seen from figs. 2 and 3 that in argon and neon ionization sets in at the atomic ionization potential for each pressure used. The spectroscopic value of the ionization potential for argon is 15·7 V, and for neon 21·5 V. The differences between these values and those given by the curves in figs. 2 and 3, amounting to 0·3 V for argon and 0·5 V for neon, are fully accounted for by contact potentials, for which no correction has been made in these figures. In argon and neon there is no evidence of ionization occurring below the atomic ionization potential even at high pressures.

The results for helium, which were taken after the argon results and before those for neon, show that ionization sets in well below the atomic ionization potential, the spectroscopic value of which is 24·5 V. Because of this the rise in the curves at the atomic ionization potential is not so sharp as in argon and neon, and consequently the correction for contact potentials to be applied to the voltage scale of fig. 4 cannot be determined as precisely as in the other two rare gases. However the rise in the three lower pressure curves appears to set in at about 23·5 V, which gives a correction of 1 V for contact potentials and energy spread in the electron beam. By subtracting the vacuum curve from the four pressure curves of fig. 4 and applying this correction of 1 V to the voltage scale we obtain the curves shown in fig. 5*a*. These curves show that ionization sets in at 19·8 V.

The two lowest excitation potentials of the helium atom are 19·77 V for the transition $1\,^1S_0 \rightarrow 2\,^3S$ and 20·55 V for $1\,^1S_0 \rightarrow 2\,^1S_0$. Now, although the voltage scale correction applied to fig. 5*a* is a little indefinite it is quite definite enough to decide between these two excitation potentials. If ionization did not occur until the atom was excited to the singlet level the voltage correction would have to be 1·8 V instead of 1·0 V. The curves of fig. 4 show that it could not possibly be as large as this, for the uncertainty in the value 1·0 V adopted is not greater than ± 0·2 V.

The conclusion we have reached is that ions are formed in helium when the bombarding electrons have sufficient energy to raise them to the first excited state above the ground state. We have now to decide whether these ions are atomic or molecular. A second collision of the excited atom with an electron, or its absorption of a photon of more than 4·8 V energy, could produce an atomic ion. However, the results for neon and argon show that the probability of this double absorption process is too small to produce ions in sufficient number to be detected. For the same reason the collision of two excited atoms, whereby one absorbs the energy of the other by a collision of the second kind resulting in an atomic ion and a normal atom, cannot occur with sufficient frequency to account for the ionization observed. Whatever the process, if the resulting ions were atomic then we should expect to obtain ions also in neon and argon at the resonance potentials of these atoms, 16·53 and 11·49 V respectively. We therefore conclude that the ions produced below the atomic ionization potential in helium are molecular ions.

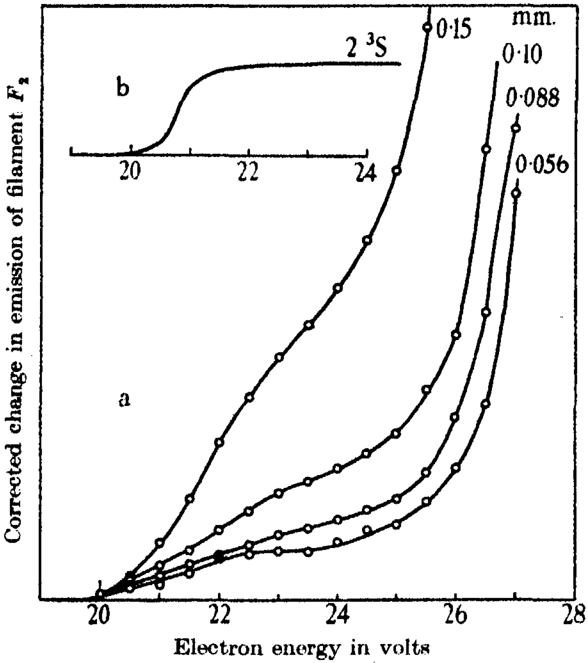


FIG. 5. (*a*) Curves of fig. 4 corrected by subtraction of the vacuum curve and with voltage scale correction for contact potentials. Showing onset of molecular ionization at the resonance potential, 19·77 V, of helium. (*b*) Theoretical excitation function for the $2\,^3S$ state of 19·77 V energy.

Helium molecules have been known since 1913 when the helium band spectrum was discovered independently by Curtis (1913) and by Goldstein (1913). During the last twenty-five years the band spectrum has been very fully investigated and has proved of supreme importance in testing the quantum theory of molecules developed by Heitler and London, Hund, Mulliken and others. A table of the energy states of the helium molecule is given by Jevons (1932). No band spectra have been observed in argon, and although a number of bands in the red in neon were reported by Dhavale (1930) this observation has not been confirmed. There is thus no definite evidence from band spectroscopy that molecules are formed in argon and neon, a negative result which is supported by our curves given in figs. 2 and 3.

Our results show that helium molecules are formed by the collision of an excited atom in the 2 ^{3}S state with either a normal atom or another excited atom in the 2 ^{3}S state. The collision with a normal atom is far more probable than a collision with a second excited atom, and, since the former process provides sufficient energy to ionize the molecule, it is probably the effective process. We thus have

$$\mathrm{He}(1s^2,\ {}^1\mathrm{S}_0) + \mathrm{He}'(1s2s,\ {}^3\mathrm{S}) \rightarrow \mathrm{He}_2^+(1s\sigma^2 2p\sigma,\ {}^2\Sigma_u) + \epsilon,$$

in which the first term represents the normal atom in its ground state 1 $^1\mathrm{S}_0$ and the second term represents the excited atom in the metastable state 2 ^{3}S of 19·77 V energy. The ionization potential of the molecule obtained from the band spectrum limit is 18·58 V (Jevons 1932), leaving 1·19 V energy to be carried away in kinetic form by the electron.

The theoretical excitation function for the metastable state occurring in the above process has been determined by Massey and Mohr (1931), and is shown in fig. 5*b*. We see that the probability of excitation to this state from the ground state is practically constant over the range of energy from 22 to 24 V. Now the lowest curve of fig. 5*a*, representing the probability of ionization at a pressure of $5{\cdot}6 \times 10^{-2}$ mm., is similar in form to the excitation function up to the atomic ionization potential, 24·5 V, and is practically flat between 22 and 24 V. The higher pressure curves in fig. 5*a* differ more in form from that of the excitation function the higher the pressure. This indicates that at low pressure the molecular ions are formed mainly from atoms in the metastable state 2 ^{3}S, while at higher pressures atoms in higher excited states also form molecular ions when they collide with normal atoms. Since these higher states are not metastable a higher pressure is required in order that an effective collision leading to attachment should occur before the atom radiates.

We have shown in our previous paper on the formation of mercury molecules (1938) that excited atoms in states other than P-states do not apparently form ionized mercury molecules by attachment even though they have more than sufficient energy to do so. The helium atom is similar to the mercury atom in that it has two outer electrons, but differs from the mercury atom in that its outer shell is complete. This difference apparently allows the formation of ionized helium molcules from excited atoms in S-states.

SUMMARY

An investigation of the formation of ionized molecules in argon, neon and helium has been made by the balanced space-charge method which had previously been used by the authors to study the formation of mercury molecules. No evidence of molecular ionization was found in argon and neon. In helium molecular ionization sets in at the resonance potential. The results show that these molecular ions are formed by the attachment of metastable atoms to normal atoms according to the process

$$\mathrm{He}(1s^2,\ {}^1\mathrm{S}_0) + \mathrm{He}'(1s2s,\ {}^3\mathrm{S}) \rightarrow \mathrm{He}_2^+(1s\sigma^2 2p\sigma,\ {}^2\Sigma_u) + \epsilon.$$

The appearance potential of the molecular ions is the energy of the 2 ^{3}S state, 19·77 V, which is 1·19 V greater than the ionization potential of the helium molecule.

Attention is drawn to the fact that these molecular ions are formed from excited atoms in S-states, whereas we have shown that mercury molecular ions are apparently formed only by excited atoms in P-states.

REFERENCES

Arnot and Milligan 1936 *Proc. Roy. Soc.* A, **153**, 359.
Arnot and M'Ewen 1938 *Proc. Roy. Soc.* A, **165**, 133.
Curtis 1913 *Proc. Roy. Soc.* A, **89**, 146.
Dhavale 1930 *Nature, Lond.*, **125**, 276.
Goldstein 1913 *Verh. dtsch. phys. Ges.* **15**, 10.
Jevons 1932 "Report on Band Spectra of Diatomic Molecules", pp. 270–1. Camb. Univ. Press.
Massey and Mohr 1931 *Proc. Roy. Soc.* A, **132**, 605.

Application of "reciprocity" to nuclei

By M. Born

(*Communicated by E. T. Whittaker, F.R.S.—Received* 15 *March* 1938)

Introduction

This paper is a continuation of a recent publication (Born 1938) under the title "A suggestion for unifying quantum theory and relativity", which I shall quote here as (I). The central question which arises from the idea of "reciprocity" is the generalization of the conception of a metric in such a way that it comprises the x- and p-space simultaneously. But as x and p have different dimensions this can be done only by making all components dimensionless. One has to divide the x-line-element by an absolute length a, the p-line-element by an absolute momentum $b = h/a$. The latter part can be neglected for all processes involving only wave-lengths long compared with a, and vice versa, corresponding to the cases of pure x-metric (relativity) or pure p-metric. If one identifies b with the limiting momentum defined in (I), it follows from (I, 22) that $a = \pi r_0$, where $r_0 = e^2/mc^2$ is the conventional radius of the electron. Instead of a and b one can consider as primary absolute constants $h = ab$ and

$$H = b/a = 0{\cdot}0085\,\text{g./sec.} \tag{1}$$

The constant H is characteristic for the idea of reciprocity in the same way as c is characteristic for relativity and h for quantum theory.

This standpoint implies that the constant b is universal. This was also assumed in the numerical calculations of (I), but no deciding argument could be given for this hypothesis.

The purpose of this paper is to support this assumption by some applications of the distribution formula (I, 8) to nuclei. Even if this formula should not be completely correct, the agreement of the calculated values of some properties of nuclei with observation can be considered as a strong evidence for the existence of an absolute momentum, or of the equivalent reciprocity constant H.

1. Nuclear radii and energies

The postulate of a closed p-space made in (I) led to the consequence that for any energetically closed system the number of quantum states is finite, namely

$$n = g\frac{\pi^2 b^3}{h^3}V = \frac{g}{\pi r_0^3}V, \tag{2}$$

where g is the weight of the state and V the spatial volume. We apply this to a nucleus of mass number $A = N + Z$ (N number of neutrons, Z number of protons). The natural assumption is that in the ground state of the nucleus each quantum state is occupied by two particles (since neutrons and protons obey the Fermi statistics, $g = 2$). Let R be the radius of the nucleus, $V = 4\pi R^3/3$. Then by putting $A = n$ with $g = 2$ we find from (2)

$$A = \frac{8}{3}\frac{R^3}{r_0^3},$$

$$R = r_0 \sqrt[3]{\tfrac{3}{8}A} = 0{\cdot}72 \times r_0 \sqrt[3]{A} = 2{\cdot}02 \times 10^{-13} \sqrt[3]{A}\ \text{cm}. \tag{3}$$

A law of this type has been assumed as the expression of the liquid drop model of the nucleus (Bethe and Bacher 1936, see especially, p. 97, § 8, formula (17)). The numerical factor has been determined by Bethe (1937) with help of the decay constants of radioactive nuclei; the value which he derives from the experimental data (formula (598), § 68, p. 167 of the report quoted) is $2{\cdot}05 \times 10^{-13}$ cm. The agreement with our value (3) is so perfect that one cannot help feeling suspicious. Therefore we proceed to the calculation of another property of the nucleus, the mean kinetic energy of its particles.

From (I, 8), (I, 9) we get for a symmetrical distribution

$$\frac{dn}{n} = \frac{dn}{A} = \frac{4}{\pi b^3}\frac{p^2\,dp}{\sqrt{(1-p^2/b^2)}}. \tag{4}$$

The average of the kinetic energy $\epsilon = p^2/2M$ is

$$\bar{\epsilon} = \int \epsilon \frac{dn}{n} = \frac{2}{\pi M b^3}\int_0^b \frac{p^4\,dp}{\sqrt{(1-p^2/b^2)}} = \frac{b^2}{2M}\frac{4}{\pi}\int_0^1 \frac{\xi^4\,d\xi}{\sqrt{(1-\xi^2)}} = \frac{3}{4}\frac{b^2}{2M} = \tfrac{3}{4}\epsilon_0, \tag{5}$$

where $$\epsilon_0 = \frac{b^2}{2M} = 2\frac{m}{M}\left(\frac{hc}{2\pi e^2}\right)^2 mc^2 = 2\frac{(137)^2}{1845}mc^2 = 10{\cdot}4 \times 10^6\ \text{e-volts} \tag{6}$$

is the maximum energy.

Very similar results have been obtained (Jordan 1937) with help of the assumption that the particles in the nucleus behave as a completely degenerated Fermi gas. Using the empirical coefficient of Bethe (mentioned above) for the radius of the nucleus Jordan finds $\epsilon_0 = 10 \times 10^6$ e-volts in excellent agreement with (6). But the two theories are essentially different, as can be seen from a comparison of the expressions for the average kinetic energy: the Fermi gas leads to $\bar{\epsilon} = \tfrac{3}{5}\epsilon_0$ instead of our formula (5), $\bar{\epsilon} = \tfrac{3}{4}\epsilon_0$. Both values are in agreement with the condition that the average binding energy U of a particle must be larger than $\epsilon_0 - \epsilon$; $U > \epsilon_0 - \bar{\epsilon}$. For one has for the

Fermi gas $\epsilon_0 - \bar{\epsilon} = \frac{2}{5}\epsilon_0 = 4 \times 10^6$ e-volts, and for our case $\epsilon_0 - \bar{\epsilon} = \frac{1}{4}\epsilon_0 = 2{\cdot}5 \times 10^6$ e-volts whereas the empirical value of the binding energy is $U = 8{\cdot}5 \times 10^6$ e-volts. One can also easily calculate the mean square fluctuation of the energy

$$\overline{\overline{\Delta\epsilon}} = \sqrt{\overline{(\Delta\epsilon)^2}}, \quad \overline{(\Delta\epsilon)^2} = \overline{\epsilon^2} - \bar{\epsilon}^2. \tag{7}$$

One finds

$$\overline{\epsilon^2} = \tfrac{5}{8}\epsilon_0, \quad \bar{\epsilon}^2 = \tfrac{9}{16}\epsilon_0, \quad \overline{(\Delta\epsilon)^2} = \tfrac{1}{16}\epsilon_0^2, \tag{8}$$

therefore

$$\overline{\overline{\Delta\epsilon}} = \tfrac{1}{4}\epsilon_0 = \epsilon_0 - \bar{\epsilon}. \tag{9}$$

The mean square fluctuation is just equal to the difference between maximum and mean energy.

2. COLLISIONS OF CHARGED PARTICLES

We have shown in (I) that the hypothesis of closed p-space leads to a modification of the Coulomb law, namely (I, 20, 23),

$$V = \frac{e_1 e_2}{r} f\left(2\frac{r}{r_0}\right), \tag{10}$$

with

$$f(x) = \int_0^x J_0(y)\,dy = \frac{2}{\pi}\int_0^1 \frac{\sin(\xi x)}{\xi}\frac{d\xi}{\sqrt{(1-\xi^2)}}. \tag{11}$$

This implies alterations in all phenomena connected with interactions of charged particles. We shall treat here the important case of collisions for which observable effects can be expected. For the sake of simplicity we assume unequal particles of high relative energy for which the cross-section is given by (see for instance Dirac 1935, p. 195, § 52, formula (15)):

$$\sigma = 4\pi^2 M^2 h^2 \,|\,(p' \,|\, V \,|\, p^0)\,|^2, \tag{12}$$

where M is the effective mass, $\vec{p}^0$ and $\vec{p}'$ the initial and final value of the relative momentum, and the matrix element

$$(p' \,|\, V \,|\, p^0) = \frac{1}{h^3}\iiint e^{i(\vec{p}_0 - \vec{p}')\vec{r}/h}\, V(r)\,dx\,dy\,dz. \tag{13}$$

If we introduce the vector which determines the change of momentum

$$\vec{P} = \vec{p}_0 - \vec{p}', \tag{14}$$

and take its direction as polar axes, $\vec{r}$ forming the angle arc cos γ with it, we have for an elastic collision

$$\vec{P}\vec{r} = Pr\gamma, \quad P = 2p_0 \sin\theta/2. \tag{15}$$

Then
$$(p' \,|\, V \,|\, p^0) = \frac{e_1 e_2}{h^3} 2\pi \int_0^\infty r\,dr \int_{-1}^{+1} d\gamma\, e^{iPr\gamma/\hbar} f\left(\frac{2r}{r_0}\right)$$
$$= \frac{e_1 e_2}{\pi h} \frac{1}{P^2} \int_0^\infty d\eta \sin\eta \,.\, f(\eta/\eta_0),$$

where
$$\eta_0 = \frac{Pr_0}{2\hbar} = \frac{p_0 r_0 \pi}{h} \sin\theta/2 = \frac{p_0}{b} \sin\theta/2. \tag{16}$$

Introducing this in (12) we get
$$\sigma = \left(\frac{M e_1 e_2}{2p_0^2 \sin^2\theta/2} f_1(\eta_0)\right)^2; \tag{17}$$

this is Rutherford's law of scattering, modified by a factor which is the square of
$$f_1(\eta_0) = \int_0^\infty d\eta \sin\eta \,.\, f\left(\frac{\eta}{\eta_0}\right) = \frac{2}{\pi} \int_0^\infty d\eta \sin\eta \int_0^1 \sin\left(\frac{\xi\eta}{\eta_0}\right) \frac{d\xi}{\xi\sqrt{(1-\xi^2)}}. \tag{18}$$

This integral can be worked out in the following way.

We consider the function
$$f_1(\eta_0, y) = \frac{2}{\pi} \int_0^y d\eta \sin\eta \int_0^1 \sin\left(\frac{\xi\eta}{\eta_0}\right) \frac{d\xi}{\xi\sqrt{(1-\xi^2)}}. \tag{19}$$

Here we can change the order of the integrations and perform first that with respect to η; then we get
$$f_1(\eta_0, y) = \frac{1}{\pi} \int_0^1 \frac{d\xi}{\xi\sqrt{(1-\xi^2)}} \left(\frac{\sin y(1-\xi/\eta_0)}{1-\xi/\eta_0} - \frac{\sin y(1+\xi/\eta_0)}{1+\xi/\eta_0}\right). \tag{20}$$

Using the formula
$$\lim_{y\to\infty} \int_a^b \frac{\sin y\alpha}{\alpha} \phi(\alpha)\, d\alpha = \begin{cases} \pi\phi(0), & \text{if } ab < 0, \\ 0, & \text{if } ab > 0, \end{cases}$$

we see that the second term in (20) tends to 0 for $y \to \infty$, whereas the first term has a limit different from 0 as long as $\eta_0 < 1$; putting $1 - \xi/\eta_0 = \alpha$ we get
$$f_1(\eta_0) = \lim_{y\to\infty} f_1(\eta_0, y) = \frac{1}{\pi} \lim_{y\to\infty} \int_{1-1/\eta_0}^1 \frac{\sin y\alpha}{\alpha} \frac{d\alpha}{(1-\alpha)\sqrt{(1-\eta_0^2[1-\alpha]^2)}} = \frac{1}{\sqrt{(1-\eta_0^2)}}. \tag{21}$$

We can now write the cross-section
$$\sigma = \sigma_0 \frac{1}{1 - \frac{p_0^2}{b^2} \sin^2\theta/2} = \sigma_0 \frac{1}{1 - \frac{\epsilon}{\epsilon_0} \sin^2\theta/2}, \tag{22}$$

where σ_0 is Rutherford's value, $\epsilon = \frac{p^2}{2M}$ the (relative) kinetic energy and $\epsilon_0 = \frac{b^2}{2M}$ the maximum energy.

The application of this formula is very much limited by the assumptions made. For an electron colliding with a proton or another nucleus one has

$$\epsilon_0 = \frac{b^2}{2m} = \frac{1}{2m}\left(\frac{h}{\pi r_0}\right)^2 = 2\left(\frac{hc}{2\pi e^2}\right)^2 mc^2 = 2(137)^2 mc^2 = 19{\cdot}200 \times 10^6 \text{ e-volts.} \tag{23}$$

But for energies of this order one cannot use these formulae, but has to apply relativistic mechanics. In any case, the effect treated here would give rise to deviations also from the relativistic cross-section for energies of at least $\epsilon_0 = 2 \times 10^{10}$ e-volts, and this result seems to be not at variance with the calculations of Bhabha and Heitler (1937).

For the proton-proton collisions the formula (22) is not directly applicable because of the exchange effect of equal particles. But one can safely assume that the deviation will begin at a characteristic energy of the order $\epsilon_0 = b^2/2M$ where M is the mass of the proton. This is precisely the energy of $10{\cdot}4 \times 10^6$ e-volts calculated in the 1st section, formula (6). Now the proton-proton collision has been experimentally investigated by Tuve, Heydenburg and Hafstad (1936) and their results have been theoretically discussed by Breit, Condon and Present (1936); the result of this careful study is that the deviations from the Coulomb scattering (Mott's formula) can be explained by an attractive interaction energy of 11×10^6 e-volts (in a distance of $2{\cdot}82 \times 10^{-13}$ cm.). The coincidence with our value is very remarkable notwithstanding the fact that the explanation of the deviations is entirely different.

For the collisions of other pairs of nuclei the simple theory does not hold because the nucleus has a finite radius over which its charge is spread. This leads to a diminishing of the deviations from the normal cross-section, with rather irregular features. It seems not worth while to study these effects in detail as long as the foundations of the theory are still unsettled.

3. Yukawa's particles

A particle moving with the velocity of light and the maximum momentum b would have the energy bc, or the mass

$$b/c = \frac{h}{\pi r_0 c} = 2\left(\frac{hc}{2\pi e^2}\right)m = 2 \times 137m = 274m. \tag{24}$$

Particles of this kind have recently been much discussed. As early as 1935, Hideki Yukawa had already suggested the existence of a new kind of field and corresponding particles for explaining the nuclear forces of small range; he estimated the mass of these particles from this range as about $30m$. Later he identified them with the penetrating particles in cosmic rays found by Anderson and Neddermeyer (1936), Street and Stevenson (1937), and Nashine, Takeuchi and Jchimya (quoted by Yukawa). The theory of these particles has been developed by Kemmer (1938), Bhabha (1938), Fröhlich and Heitler (1938), and Yukawa himself (1938). The values of the masses given by these authors differ appreciably. Street and Stevenson (1937) find $120m$, Fröhlich and Heitler $80m$, the Japanese investigators $\frac{1}{10}$ proton mass or about $180m$. Our value (24) is of this order. As stated in I, our numerical results are preliminary as only the electrostatic part of the self-energy of the electron was used for the determination of b.

SUMMARY

The formula for the distribution of quantum states, which follows from the principle of reciprocity and the assumption of a closed p-space, is applied to some properties of nuclei. The results can be considered as confirmations of the hypothesis. The mass of a particle moving with velocity of light and maximum momentum is of the same order as that of the particles which Yukawa has introduced.

REFERENCES

Anderson and Neddermeyer 1936 *Phys. Rev.* **50**, 273; **51**, 884.
Bethe 1937 *Rev. Mod. Phys.* **9**, 69.
Bethe and Bacher 1936 *Rev. Mod. Phys.* **8**, 82.
Bhabha 1938 *Nature, Lond.*, **141**, 117.
Bhabha and Heitler 1937 *Proc. Roy. Soc.* A, **159**, 432.
Born 1938 *Proc. Roy. Soc.* A, **165**, 291.
— 1938 *Nature, Lond.*, **141**, 327.
Breit, Condon and Present 1936 *Phys. Rev.* **50**, 825.
Dirac 1935 "Quantum Mechanics", 2nd ed. Oxford Univ. Press.
Fröhlich and Heitler 1938 *Nature, Lond.*, **141**, 37.
Jordan 1937 *Ergebn. Exakt. Naturw.* **16**, 47.
Kemmer 1938 *Nature, Lond.*, **141**, 116.
Street and Stevenson 1937 *Phys. Rev.* **51**, 1005; **52**, 1003.
Tuve, Heydenburg and Hafstad 1936 *Phys. Rev.* **50**, 806.
Yukawa 1935 *Proc. Phys.-Math. Soc. Japan*, **17**, 48.
— 1937 *Proc. Phys.-Math. Soc. Japan*, **19**, 712.
— 1938 *Proc. Phys.-Math. Soc. Japan*, **20** (typewritten, separate pagination).
Yukawa and Sakata 1937 *Proc. Phys.-Math. Soc. Japan*, **19**, 1084.

The adsorption of vapours at plane surfaces of mica

II. Heats of adsorption and the structure of multimolecular films

By D. H. Bangham and S. Mosallam

(*Communicated by D. L. Chapman, F.R.S.—Received* 25 *March* 1938)

In a previous paper, where the results of adsorption measurements with known surface areas of mica were recorded (Bangham and Mosallam 1938), attention was focused on the range where the first molecular layer is yet incomplete. The present communication deals mainly with the transition to a thicker covering as the pressure of vapour is increased towards saturation. Conclusive evidence has already been given that the very large increase in the adsorption of benzene and carbon tetrachloride near saturation is not attributable to the condensation of liquid, either between contiguous faces of the mica strips (in those experiments where these were not separated by wires) or in incipient cleavages at their edges. The adsorbed films were found incapable of acting as nuclei of condensation for the bulk liquids, and were proved to constitute a phase thermodynamically distinct from the latter. Nevertheless, as the sequel will show, the adsorption energy is very little different from the normal heat of condensation once the first monolayer is complete, and there is evidence which suggests that the densities of the multimolecular films may not differ widely from those of the bulk liquids.

EXPERIMENTS WITH BENZENE VAPOUR

In eight experiments with this vapour, using the adsorption vessel A,* the observational points on the isotherms were fairly widely spaced, and it was possible in most cases to draw a smooth sigmoid curve which left only one or two of these outside the allowable margin of error. Attempts to determine accurately the adsorption value at the point of inflexion were frustrated, however, by the fact that the results, though fairly reproducible, were not entirely so; and also on account of the persistent occurrence of apparently irregularly placed points precisely in this region.

---

* In this vessel strips of mica of aggregate area 18,000 cm.^{2} were placed face to face, without separating wires.

Several adsorption and desorption experiments were carried out to discover whether these irregularities were attributable to any fault in the technique of measurement. The desorption experiments showed the isotherm to be reversible if the pressure was reduced before completion of the first molecular layer, that is to say, in the region where irregularities were absent. But if the adsorption had been carried beyond this point, the desorption curve showed a definite hysteresis effect persisting down to very low pressures (fig. 1). Since the quantities of vapour ultimately recovered (by freezing part of the measuring system and warming the mica tube) were in all cases in good agreement with those calculated to be present, there is reason to believe that this hysteresis was real, and not due to inaccuracy of measurement.* Nor was it due to insufficient time being allowed the system to come to equilibrium. In general, adsorption measurements made shortly after an admission or removal of vapour differed by no more than a few mm.^{3} at N.T.P. from readings taken after standing for one or several days.

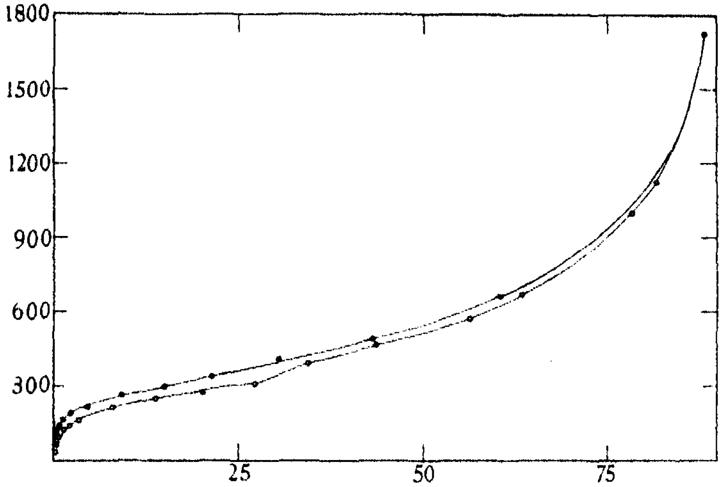


FIG. 1. Adsorption ○ and desorption ● isothermals of benzene. Abscissae, pressures; ordinates, adsorption in mm.^{3}. Area of mica, 18,000 cm.^{2}.

Heat of adsorption. In one of the desorption experiments with tube A the temperature of the thermostat was changed from 25 to 35° C. at each pressure reduction, 24 hr. being allowed for the system to attain equilibrium before reading. The former temperature was then restored, and, after taking a further reading as a test of reversibility, the next pressure reduction was made. Since, on the original temperature being restored, the adsorption invariably returned to very near its original value (often to within 1 mm.^{3} N.T.P. of this), it is certain that the equilibrium was here reversible with respect to the small changes of temperature and pressure involved, so that a calculation of the differential heats of adsorption from these data with the aid of the usual thermodynamic equation is fully justified.*

---

* In experiments where the pressure has been raised to near saturation, a much larger margin of error must of course be allowed for, than in those where the pressure was kept low.

The results, shown as ordinates in fig. 2 against the areas per molecule as abscissae, are of considerable interest. The adsorption energy is very nearly constant for small coverings, its value, where the area per molecule exceeds 35 A^2, being about 12,000 cal./mol. As the area per molecule decreases from 35 to *ca.* 25 A^2 the energy falls rapidly to a minimum value; thereafter there is a slight rise to a value very near the normal heat of liquefaction.

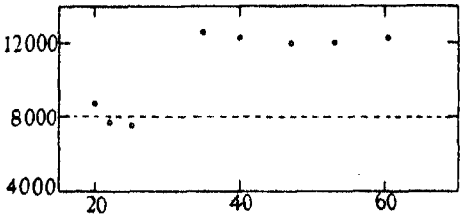


FIG. 2. Abscissae, areas per molecule in Angström units; ordinates, adsorption energy in calories per molecule. The broken line indicates the latent heat of the normal liquid.

Discontinuities in isothermals. When it was at length realized that the irregularities in the benzene isotherms—amounting, in some cases, to discrepancies of about 50 mm.^{3} N.T.P., or ten times the allowable margin of error—were not due to the onset of capillary condensation, an attempt was made to trace their course in greater detail by taking a greater number of readings. Where possible, the compression ratio used in reading the gauge was varied, so that each point on the resulting isotherm should represent not a single observation but the mean of several made under slightly different conditions.

---

* The procedure adopted was as follows: Two plots of the isotherms were made, one of the adsorption against log p (p = pressure), and the other of the logarithm of the adsorption against this variable. Smooth curves were then drawn through the points, and on each graph the intercepts parallel to the log p axis were read off. The two methods of plotting gave results in excellent agreement, and the means of the two have been used in fig. 2. In drawing the curves it was necessary to ignore one point on the 35° isotherm. *Assuming* this point to have been widely in error, the remainder would have been unaffected, for no vapour was removed from the system at this stage. The inference to be drawn from the heat of adsorption curve in fig. 2 is, therefore, that even when the experimental isotherms are "smoothed out", there yet remains evidence of a fairly abrupt change in the nature of the adsorption when the area per molecule increases from 25 to 35 A^2. If, as is highly probable, the abnormal point was *not* in error, the discontinuity in the energy curve would be even more marked.

Fig. 3 shows the results of two separate experiments carried out in this way at 25 and 35° C. respectively. For these the adsorption vessel *B* was

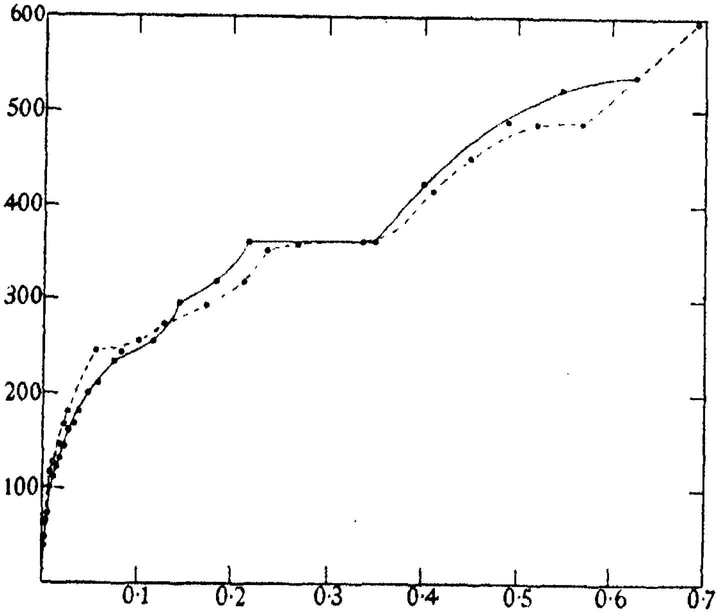


FIG. 3. Isothermals of benzene. Abscissae, relative pressures; ordinates, adsorption in mm.^{3} N.T.P. White circles, 25° C.; black circles, 35° C. Area of mica, 19,000 cm.^{2}.

used, containing mica strips of aggregate area 19,000 cm.^{2}, separated by wires. It will be observed that in the two graphs (which are plotted on a natural scale, but with the *reduced* pressures as abscissae) the irregularities take very much the same form. No doubt can remain as to the complex nature of the isotherms. The discontinuities (or sharp changes of curvature) recall those found by Allmand and his co-workers (1931) with charcoal and silica gel, but are more widely spaced. Most conspicuous are the horizontal sections of the isotherms near the middle of the figure (area per molecule 19·7 A^2), occurring between two regions of opposite curvature, both of which run counter to the general trend of the graph as a whole. (The cross-sectional area of the benzene molecule perpendicular to the plane of the ring is probably near 21·5 A^2.)

These two graphs also confirm the conclusions already drawn regarding the adsorption energy. The abscissae being fractions of the saturation pressures, it follows that where, for a given adsorption, the 25° points lie to the *left* of the 35° C. curve, the differential heat of adsorption must *exceed* the normal heat of liquefaction; where they lie to the right the opposite will be the case. It will be observed that the relative positions of the curves become reversed when about 275 mm.^{3} have been adsorbed, the area per molecule* being then 25·8 A^2, and the pressure about one-seventh of saturation.

Had the graphs of fig. 3 been plotted with a logarithmic abscissa scale, the horizontal intercepts between the curves for pressures $> 0{\cdot}3$ of saturation would appear very small; and since a much larger margin of error must be allowed when adsorption heats are calculated from separate experiments than when the temperature is varied without addition or withdrawal of vapour, the conclusion to be drawn is that above this pressure (at which the area per molecule is 19·7 A^2) the heat of adsorption differs from the ordinary latent heat by no more than the experimental error.†

EXPERIMENTS WITH METHYL ALCOHOL

The adsorption vessel *C* was used; this contained mica strips of aggregate area 24,000 cm.^{2}, separated by fine wires. Adsorption experiments only were attempted.

Heat of adsorption. In each of the runs the temperature of the thermostat was changed at each step from 35° to 25° C. and back again, to obtain data for calculating the adsorption energy. The results are graphed in fig. 4, where the common logarithm of the relative pressure p/p_0 (p_0 = saturation pressure) is plotted as abscissa, and the adsorption as ordinate.

---

* The apparent agreement here with the value of 25·4 A^2 found by Adam (1930) for the effective cross-section of the tilted rings of benzene compounds at the surface of water is almost certainly fortuitous; but we regard it as established that both the adsorption energy and the slope of the isotherm undergo an abrupt change at *about* the adsorption value to be expected for a unimolecular film of molecules oriented on edge.

† The intercepts between the curves of fig. 3 in the region of *low* pressures, though small, represent, of course, an energy change which is considerable; a natural, in preference to a logarithmic scale has been chosen because it shows more clearly the magnitude of the irregularities and their relation to possible errors of measurement.

The isotherms are clearly continuous until the pressure reaches about 0·04 of saturation. The adsorption energy is also constant until the area per molecule falls below *ca.* 17 A^2; this is shown by the constancy of the horizontal intercepts between the two curves. To the value of 3900 cal./mol. calculated from the intercept must be added 9500 cal. for the normal latent heat, making in all 13,400 cal./mol. for the constant adsorption energy in the first molecular layer (area per molecule > 17 A^2).

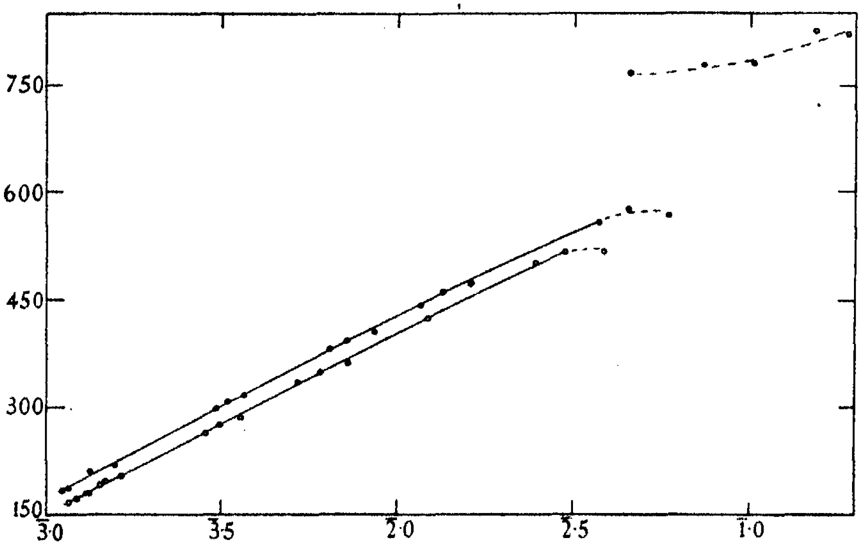


Fig. 4. Isothermals of methyl alcohol. Abscissae, values of $\log_{10} p/p_0$ (p/p_0 = relative pressure); ordinates, adsorption in mm.^{3} on 24,000 cm.^{2} of mica. White circles, 35° C.; black circles, 25° C.

Above 0·04 of the saturation pressure discontinuities appear, the adsorption being increased 60 % or so by a small increase of pressure. Simultaneously, the adsorption energy decreases abruptly to a value probably not very different from the normal latent heat; for though the experimental points above the break (a few of which are plotted in the figure) are too irregular to permit an accurate estimate of the adsorption energy in this region, those obtained at 25° C. lie systematically neither to the one side nor the other of those at the higher temperature.

It can safely be concluded that the break is connected with a sudden thickening of the film, probably by formation of a second molecular layer.

EXPERIMENTS WITH CARBON TETRACHLORIDE

As stated in Part I, the isotherm of this substance appears to consist of a series of loops, each of which is concave to the adsorption axis (fig. 5 of this paper); further observations are, however, desirable, and the heat of adsorption was not measured.

DISCUSSION

It is illuminating to combine the isotherms of all three experimental vapours, plotting the fractions of the saturation pressures as one variable. This has been done in fig. 5, where the ordinates represent the thicknesses

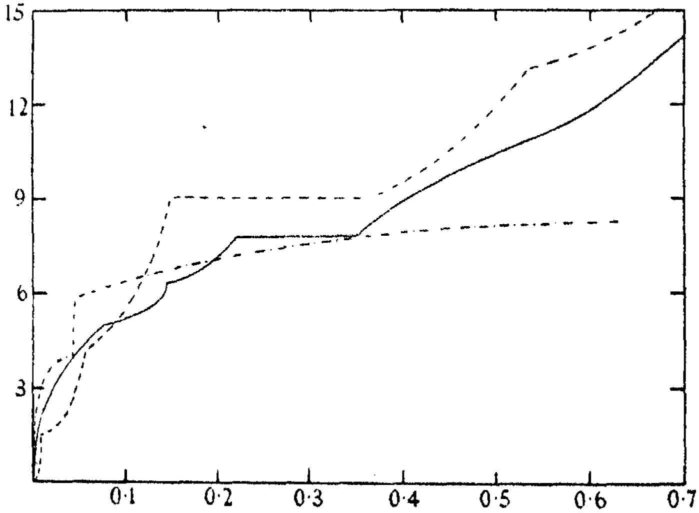


FIG. 5. Apparent thicknesses of films in Angstrom units (ordinates) at different relative pressures (abscissae). Benzene ——— ; carbon tetrachloride - - - -; methyl alcohol -·-·-·-·-. The densities of the films are assumed equal to those of the normal liquids.

(in Angstroms) *which the films would have if their densities were equal to those of the normal liquids*. Thus represented, the isotherms—whether fortuitously or otherwise—are seen to lie surprisingly close together over much of the range. Differences of molecular architecture appear to assert themselves most markedly in the finer details of the isotherms, and in the low pressure relations which the use of a natural scale tends to obscure.

This manner of plotting the data was suggested by the results of experiments with "visible" films such as are produced by exposing new mica surfaces to jets of air charged with the supersaturated vapours (Bangham, Mosallam and Saweris 1937). Elsewhere it will be shown that these thick films (which give rise to interference colours from which their thicknesses can be roughly gauged) have many properties different from the bulk liquids, but *identical with those of the films formed by the unsaturated vapours* with which we are here concerned. It was observed that films of approximately the same thickness were formed from vapours which must necessarily have been supersaturated to about the same extent.* The surprising fact that a relation of this type applies roughly to the *fictitious* film thicknesses plotted in fig. 5 and to pressures as low as 0·05 of saturation, would suggest that the densities even of these very thin films may be not widely different from those of the bulk liquids. Apart from this the graph is of interest as an example of an approximate "equal volume" relationship which cannot possibly depend upon the filling up of a fixed pore space.†

The present experiments do much to reconcile Langmuir's conception of unimolecular adsorption (in its modified form where the molecules are considered mobile) with the older hypothesis of a multimolecular film of which the thickness is determined by the degree of saturation of the vapour. The approach to completion of the Langmuir monolayer has been shown to be accompanied by discontinuities (or sharp changes of curvature) in the isotherms, and by a more or less abrupt decrease in the adsorption energy to a value near the normal heat of liquefaction. But although the thicker films have much the same energies, and possibly also much the same densities as the bulk liquids, the states of aggregation in the adsorbed and in the liquid phases cannot be quite the same; for the angle of contact is not zero when the vapours are saturated,* nor even when there is sufficient supersaturation to render the films visible. An adequate explanation of this must clearly await new knowledge as to liquid structure.

---

* Since methyl alcohol behaves no differently from benzene and carbon tetrachloride, there can be no doubt that the isotherm of this substance turns away from the pressure axis (like those of the other substances named) beyond the experimental region.

[† *Note added in proof.* Mention should be made here of the recently published measurements by Palmer and Clark (1935) and Palmer (1937) of the adsorption of vapours at known surfaces of vitreous silica. Palmer finds that the isotherms of benzene, acetone and methyl alcohol are asymptotic to axes at $1{\cdot}021\,p_0$, $1{\cdot}016\,p_0$ and $1{\cdot}014\,p_0$ respectively; the adsorption at the saturation pressures p_0 are therefore finite, in accordance with the present authors' findings. He also infers from the form of the empirical equation which reproduces the data that the rapid increase of adsorption near saturation is not due to capillary condensation, even though his surfaces were "rough". Palmer's isotherms are represented for the most part by smooth sigmoid curves of the type shown in fig. 1 of this paper, but it must be borne in mind that his adsorbing surface was only about one-quarter of that used by us, so that such irregularities as we found in the benzene isotherms, for example, would almost certainly escape notice. Palmer and Clark's low pressure data for cyclopentane show some points of similarity to our measurements for carbon tetrachloride: there is a sharp change of curvature (if not an actual discontinuity) when the adsorption is only a fraction of a unimolecular layer. Palmer suggests that *equal numbers of molecular layers* (of different substances) are adsorbed at equal relative pressures; but his calculations of the quantities in the monolayers, made with the aid of an empirical variation of Langmuir's equation, appear to us to be open to some doubt. Probably of greater significance is his result that the saturation adsorptions of six different substances are approximately equimolecular; and it may be noted that the three graphs of our fig. 5 are tending, at higher pressures, to relative positions which would conform roughly with this rule. Should the rule in question be at all general, the fact that our three experimental substances form films of nearly equal thickness between 0·1 and 0·4 of the saturation pressure must almost certainly be fortuitous.]

The different condensed phases of a substance do not differ widely in density, and recent well-supported theories of the liquid state have assumed the same forces to be operative there as in the crystal, and the same orientations to be imposed, but with a degree of randomness which is foreign to the crystalline state. It is but reasonable to suppose that these orientations will tend to reappear, modified no doubt by the force field of the adsorbent, in the adsorbed films formed from the vapours; such an hypothesis would not necessarily imply their identity with the bulk liquids.

Adapting the method of Prins (1929) and others, let us write $g(r)$ for the density of molecular centres at distance r from the surface. If we neglect the small concentration in the free-vapour phase, the total number of molecules adsorbed is

$$\int_0^\infty g(r)\,dr.$$

Langmuir's theory requires that $g(r)$ should rise sharply to a single peak (curve A, fig. 6). According to the older "potential" theory of Polanyi (1920) the curve should be nearly flat for distances outside the molecular radius, until the limit of the supposedly liquid condensed phase is reached, where there should be an abrupt fall (graph B).

Considerations of molecular size and arrangement oblige us to suppose the correct form of $g(r)$ to be given by some such curve as C in the diagram. Thus accepting Polanyi's view that the forces *mainly* instrumental in bringing about the adsorption of *vapours* are the "dispersion" forces which play so important a part in liquefaction, but rejecting his assumption that the films are liquid and molecularly chaotic, it becomes possible to discern why the isotherms of different substances should be so similar in general trend, yet so very diverse in point of detail. The adsorption energy of a molecule, since it will depend upon the number of its near neighbours in the film, will be greater when a layer is nearing completion than when a new layer is begun.

---

* The statement, appearing in Part I, that the equilibrium contact angle is less than 90° and much greater than 0° should be amended to read "much less than 90° and greater than 0°".

The visualization of the film structure in terms of a distribution function leads us to a restatement, in more general and accurate terms, of a suggestion made by Semenoff (1930) that the completion of each adsorbed layer should be associated with a break in the isothermal. A bending away from the adsorption axis is liable to occur whenever the film thickness corresponds to a well-defined molecular spacing in the bulk solid or liquid phases of the adsorbed substance; a bend away from the pressure axis should occur whenever a new layer is nearing completion, or a reorientation is in progress. But, given molecules of any degree of complexity, we should not expect the film thickness necessarily to increase by integral multiples of a single parameter, for reorientation might well play an important part.

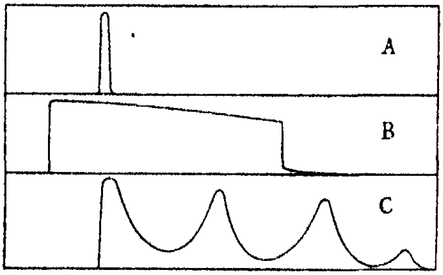


FIG. 6. Distribution of molecular centres in the neighbourhood of a surface; *A*, according to Langmuir's theory; *B*, according to Polanyi's older hypothesis of a thick film; *C*, according to hypothesis of laminated multimolecular film.

With any but the simplest molecular species we should expect the orientation of the innermost layer or layers to become affected as the adsorption increases, and this is a very probable cause of the hysteresis observed in these and other experiments. It is significant that, with benzene, no hysteresis was observed if the pressure reversal was made before the first breakpoint was reached. Initially, at all events, the first monolayer appears to be differently packed from the succeeding ones, the orientation often being such as to diminish the effects of the intermolecular attractive forces.*

* Many authors consider that repulsive forces of long range are set up between similarly oriented permanent or induced dipoles; e.g. Magnus (1929), J. K. Roberts (1935). This explanation of the "two-dimensional gas" behaviour of some films can

Thus the isotherms of benzene and methyl alcohol are smoothly concave to the pressure axis for small coverings, the differential heat of adsorption being constant (or nearly so) and markedly greater than the normal heat of liquefaction.

With some adsorbents the packing in the Langmuir layer must be such that even near saturation there is little or no tendency for the films to thicken. At all events much support has been found for the view that the films formed even from saturated vapours are of no more than molecular thickness.*

Comparison with X-ray data. A comparison of the graphs of fig. 5 with the molecular spacings found by X-ray methods is rendered difficult, and perhaps unprofitable, by the absence of precise information as to the densities of the films. Attention may, however, be drawn to the break in the benzene isothermal where the film thickness is *ca.* 5 A; liquid benzene gives a characteristic diffraction maximum at 4·7 A (Stewart 1929). With methyl alcohol there is a sharp turn away from the adsorption axis at *ca.* 4 A, which may also correspond to the spacing in the liquid of 3·8 A found by Stewart and Morrow (1927); a film of about twice this thickness persists over a very wide range of pressures. The importance of these possibly fortuitous cases of agreement should not, however, be overestimated.

Intermolecular forces. To account for the existence of thick adsorbed films, De Boer and Zwikker (1929), and more recently Bradley (1936*a, b*), have suggested a propagation of induced polarization of the molecules from layer to layer. Though calculation predicts (Lennard-Jones and Dent 1928) and experiment confirms (Bradley 1936*a, b*) that the deforming power of potassium ions (with which we are here concerned) is rather small, the theory has attractive possibilities, and accounts well for the fact that many adsorbents, including charcoal (which we regard as essentially non-polar), do not give rise to thick films even at saturation.

---

hardly apply where the heat of adsorption is as constant as it is with methyl alcohol. With benzene, also, no evidence was found of a marked diminution of adsorption energy in the range of formation of the Langmuir layer; but the accurate data of fig. 3 refer, actually, to a desorption experiment, so that the arrangement of molecules in the monolayer may have been different from that obtaining during adsorption (*vide* text).

* See, for example, Frazer, Patrick and Smith (1927); these authors ascribe the very large adsorptions on glass found by McHaffie and Lenher (1925) to the roughening of the surface and the formation of a gel-like layer of silica. Using a highly sensitive optical method of measurement, J. H. Frazer (1929) was able to detect a marked thickening of the film formed by methyl alcohol on glass at about 90 % saturation pressure, but no condensation of toluene, acetone, etc., even at of saturation.

The equations developed for the isotherms are

$$\log p/k_3 p_0 = k_2 k^n \quad \text{(De Boer and Zwikker)}, \tag{1}$$

where $k_1 < 1$, k_1, etc., are constants, and n is the number of molecular layers adsorbed at pressure p; and

$$\log p_0/p = k_1 k_3^q + k_4 \quad \text{(Bradley)}, \tag{2}$$

where $k_3 < 1$, k_1, etc., are constants, and q is the adsorption. Equations of this type may be said, in general terms, to express the expectation that the "adsorption potential" $RT \log_e p_0/p$ will fall off according to a compound interest law as the dipole induction effect becomes weaker in successive layers. This statement is not quite accurate, however, in the case of De Boer and Zwikker's equation, for this has the special merit of predicting a *finite* adsorption at saturation: the curve becomes asymptotic to an axis at $p > p_0$, thus allowing for the possible existence of supersaturation effects in the vapour phase, and of a finite equilibrium angle of contact with the liquid.

Since most of our data for mica refer to films within the effective range of the "dispersion" forces of the solid (which, when summed for all the atoms in a plane, fall away much less rapidly than the inverse sixth power of the distance, this being the power law applying to individual atoms), no attempt at comparison with the theoretical equations will be made here. The constant heats of adsorption found by us for benzene and methyl alcohol in the unimolecular films would suggest the absence of long-ranged tangentially directed repulsive forces such as would result from the interaction of dipoles similarly oriented with their axes normal to the surface.

Effect on surface tension of the adsorbent. Of interest in connexion with the interplay of the forces of adhesion and cohesion are the calculations of the surface energy lowering of the adsorbent, made with the aid of Gibbs's adsorption equation. In its integrated form this relation becomes

$$F = RT \int_{p=0}^{p=p} \Gamma d \log_e p, \tag{3}$$

where Γ is the surface excess concentration in mol./cm.^{2}, and F the surface energy lowering at pressure p. Provided equilibrium is maintained between the film and the vapour phase, the use of this equation over the range of a discontinuity is justified; and, since we are dealing with mobile films, F can rightly be regarded as defining a two-dimensional "pressure" which diminishes the surface tension of the mica.* Using a graphical method of integration, the following values of F have been calculated for different relative pressures of benzene vapour. For comparison, the values which F would have if the film remained unimolecular are also inserted in Table I.

---

* For a discussion of these points see Bangham (1937).

TABLE I. DECREASE OF SURFACE TENSION DUE TO DIFFERENT RELATIVE PRESSURES OF BENZENE VAPOUR AT 25° C.

Relative pressure	0·057*	0·1	0·25	0·5	0·75	0·94	1·0
Surface tension decrease (dynes/cm.)	33·7	40	57	73·5	86	99	(108)†
Decrease due to Langmuir monolayer	33·7	—	55	64·5	70·5	73·5	75

* Corresponds to first breakpoint. † Extrapolated.

The graph of F against the relative pressure (not shown) has relatively little curvature over the experimental range, and extrapolation over the small remaining interval up to full saturation is probably justified, for the saturation adsorption must necessarily be finite with a finite angle of contact. It will be seen that though the film thickens so markedly near saturation, the greater part of the surface tension lowering is still due to the Langmuir monolayer. This is true even of the lowering produced on immersion of the mica in liquid benzene: for if F_S is the surface tension lowering caused by the saturated vapour and F_L that caused by the liquid, we have

$$F_L = F_S + \gamma_{LV} \cos\theta, \tag{4}$$

θ being the contact angle and γ_{LV} the surface tension of the liquid. Since θ, though finite, is quite small, a sufficiently close approximation is given by adding together F_S and γ_{LV}. The resulting value of F_L at 25° C. is 136 dynes/cm.

The data for carbon tetrachloride and methyl alcohol are too incomplete to permit a near approximation to be made to the values of F_S; a rough calculation gives a value near 100 dynes/cm. for the former vapour, and near 250 dynes/cm. for the latter; the latter figure is to be regarded as a lower limit, for the effect due to any possible further thickening of the film just below saturation has been neglected. It will be noted that the F_S values for the three substances stand in the *inverse* order of the apparent film thicknesses (fig. 5) at pressures approaching saturation.

One of the authors (D. H. B.) is grateful to the Leverhulme Fellowship Trust for the award of a Fellowship. The experimental work described was carried out at the Egyptian University, Cairo, during the years 1931–7.

SUMMARY

The results are given of measurements of the quantities of benzene, methyl alcohol, and carbon tetrachloride adsorbed at a known surface area of mica at pressures approaching saturation. The adsorption energy per mol. is constant whilst the first molecular layer is being formed (benzene, methyl alcohol), but its completion is marked by an abrupt decrease to a value near the normal heat of liquefaction. The graphs of apparent film thickness against relative pressure for the three substances are not widely divergent over much of the range investigated, but the isotherms show discontinuities or sharp changes of curvature at adsorption values which depend on the specific properties of the molecules, and are thought to be related to molecular spacings in the bulk condensed phases. Approximate values are given for the surface tension lowering of the mica, caused by the saturated vapours.

REFERENCES

Adam 1930 "The physics and chemistry of surfaces", p. 96.
Allmand and Burrage 1931 *Proc. Roy. Soc.* A, **130**, 610.
Bangham 1937 *Trans. Faraday Soc.* **35**, 805.
Bangham and Mosallam 1938 *Proc. Roy. Soc.* A, **165**, 552.
Bangham, Mosallam and Saweris 1937 *Nature, Lond.*, **140**, 237.
Bradley 1936*a* *J. Chem. Soc.* p. 1467.
— 1936*b* *J. Chem. Soc.* p. 1799.
De Boer and Zwikker 1929 *Z. phys. Chem.* **3**, 407.
Frazer 1929 *Phys. Rev.* **33**, 97.
Frazer, Patrick and Smith 1927 *J. Phys. Chem.* **31**, 897.
Lennard-Jones and Dent 1928 *Trans. Faraday Soc.* **24**, 93.
McHaffie and Lenher 1925 *J. Chem. Soc.* **127**, 1559.
Magnus 1929 *Z. phys. Chem.* **142**, 401.
Palmer 1937 *Proc. Roy. Soc.* A, **160**, 254.
Palmer and Clark 1935 *Proc. Roy. Soc.* A, **149**, 360.
Polanyi 1920 *Z. Electrochem.* **26**, 370.
Prins 1929 *Z. Phys.* **56**, 617.
Roberts 1935 *Proc. Roy. Soc.* A, **152**, 445.
Semenoff 1930 *Z. Phys. Chem.* **7**, 471.
Stewart 1929 *Phys. Rev.* **33**, 889.
Stewart and Morrow 1927 *Phys. Rev.* **33**, 232.

The swelling of charcoal

Part V. The saturation and immersion expansions and the heat of wetting

BY D. H. BANGHAM AND R. I. RAZOUK

(*Communicated by D. L. Chapman, F.R.S.—Received* 25 *March* 1938)

INTRODUCTION

In previous papers of this series evidence was advanced that:

(i) The expansion which unactivated wood charcoal undergoes when adsorbing gases or vapours is directly proportional to its surface (free) energy lowering (Parts II and III*).

(ii) The Gibbs adsorption equation, relating the adsorption, the surface energy decrement of the charcoal, and the pressure of the gas, is valid except when a change of surface phase is in progress (Parts II and III).

(iii) The films of the lower alcohols adsorbed on charcoal obey very much the same two-dimensional equations of state as they do when adsorbed on mercury (Parts III and IV*).

These results we shall refer to as relationships (i), (ii) and (iii) respectively.

The implications of (i) and (ii) are so fully borne out by experiment that they need no longer be questioned. The third, though very well supported by experiment, and leading to twelve concordant values of the adsorbing area of the charcoal used which agree with independent estimates, stands on rather a different footing. The special point of doubt relates to whether the imperfections of the charcoal surface can be so far overlooked that the films, like those on mercury, can be regarded as bounded only on one side by the adsorbing surface. In the present paper (iii) is regarded as still *sub judice*; and though the evidence to be given will lead us to decide in its favour, it is desirable to make a clear distinction between those of our results which depend on its validity and those which do not.

Let us suppose s grams of a gas or vapour, of molecular weight M, to be adsorbed on 1 g. of charcoal, of which the effective adsorbing area (specific surface) is Σ; let x be the linear expansion per cent, F the surface (free) energy decrement in ergs/cm.^{2}, and p the pressure of the gas or vapour. Relation (i) gives us

---

* Part II, Bangham, Fakhoury and Mohamed 1932; Part III, Bangham, Fakhoury and Mohamed 1934; Part IV, Bangham 1934.

$$x = \lambda F, \tag{1}$$

where λ is a constant, probably related to the elastic constants of the charcoal.

Relation (ii), assuming ideal behaviour on the part of the free gas or vapour, may be put into the form

$$\frac{s}{m\Sigma} = \frac{dF}{RTd\log_e p}, \tag{2}$$

where $s/M\Sigma$ is the adsorption in moles/cm.^{2}, and R and T have their usual meanings. Combining equations (1) and (2) we have

$$\frac{\lambda}{\Sigma} = \frac{Mdx}{sRTd\log_e p}, \tag{3}$$

in which all the terms on the right-hand side are experimentally determinable, so that the quotient λ/Σ can be readily calculated from expansion-adsorption-pressure measurements. This ratio is known to be independent of the temperature,* at all events over a limited range.

The less firmly established relation (iii) enables the separate evaluation of λ and Σ; with its aid it becomes possible to calculate the surface energy decrements corresponding to observed expansions, the thickness of the films, and the effective cross-sectional areas of the molecules.

The present investigation deals mainly with the behaviour of charcoal on *wetting*. Charcoals are notoriously dissimilar in their behaviour, and this is very often due entirely to differences of specific surface. Could the energy changes accompanying the contact of 1 cm.^{2} of adsorbing surface with a given liquid once be established, a simple determination of the heat of wetting would suffice, in favourable cases, for the determination of Σ (cf. Harkins and Ewing 1921; Patrick and Grim 1921; Bartell and Fu 1930). Relation (iii) affords the means for doing this.

Preliminary measurements of the heat of wetting of wood charcoal by methyl alcohol gave consistently higher results than those derived by the Gibbs-Helmholtz equation from the calculated surface energy decrements at *saturation* and their temperature coefficients, though the same charcoal had been used both for the extensometric and for the calorimetric determinations. The calculated values (*ca.* 13 cal./g. as against 16–17 cal./g.

---

* Bangham, Fakhoury and Mohamed (1934); further evidence is given in the present paper.

found experimentally) do *not* depend upon the value assigned to Σ; for, if x_s and F_s are the saturation expansion and surface free energy decrement respectively, we have, using the Gibbs-Helmholtz equation in conjunction with (1), and assuming λ to be independent of the temperature,

$$\text{*Heat evolved per g. of charcoal} = \Sigma\left(F_s - T\frac{dF_s}{dT}\right) \tag{4}$$

$$= \frac{\Sigma}{\lambda}\left(x_s - T\frac{dx_s}{dT}\right). \tag{5}$$

The ratio λ/Σ is accurately known, and its value in no way depends upon the validity of relation (iii).

Three possible causes suggest themselves for the small but significant difference between the observed and calculated heat effects:

(*a*) In the range of very large expansions the extensometer previously used may have given persistently low results.

(*b*) The proportionality factor λ may increase with the temperature; if this is so, equation (5) must be rewritten

$$\text{Heat evolved per g. of charcoal} = \frac{\Sigma}{\lambda}\left(x_s - \frac{T\,dx_s}{dT} + \frac{Tx_s}{\lambda}\frac{d\lambda}{dT}\right). \tag{6}$$

(*c*) The charcoal may reach a different equilibrium when exposed to the saturated vapour from what it does when immersed in the liquid.

Regarding (*a*) it was known that the recorded values of the saturation expansions, lying as they did near the limit of movement of the extensometer, were liable to be rather too small; the error had been roughly estimated, however (Part II), and found to be of a smaller order than would account for the discrepancy.

A variation of λ with temperature (*b*) is rendered most unlikely by the fact that the ratio λ/Σ is found experimentally to be invariant. Moreover, in those cases (e.g. with methyl alcohol) where theory predicts that the surface energy decrement for a given adsorption should be directly proportional to the absolute temperature, an identical relation is found experimentally for the expansion (Bangham 1937).

---

* This heat is not the total heat of adsorption from the vapour, but the heat of adsorption *less* the normal latent heat of sufficient vapour to saturate the charcoal. Williams (1918) makes this point very clear. Whilst, however, Williams considered that the above expression should give the heat of *wetting*, it is here shown that it gives only the heat of *saturation*, which is different from the heat of wetting (cf. Bangham and Razouk 1937).

The experiments to be described show (*c*) to be the main cause of the difference between the calculated and observed heat effects. It is shown that charcoal, already expanded to the maximum extent by the saturated vapour of methyl alcohol, swells still further when immersed in the liquid; and that equation (1) applies to the further lowering of surface energy which takes place when the charcoal surface becomes *wetted*.

Other theories of the expansion effect. That the saturation expansion x_s should differ from the immersional value x_L is contrary to other current theories of the effect, and it also eliminates certain otherwise possible models for the structure of charcoal.

The hygrometric properties of hair, and the volume changes of concrete and other building materials with atmospheric moisture are commonly ascribed to variations of hydrostatic pressure in interstitial layers of *liquid* water which is assumed to be held by capillary condensation (Whipple 1922, Freysinnet 1936). As the relative humidity increases, the water/air meniscuses become less concave, and the *negative* pressure within the interstitial water layers is correspondingly diminished. Since liquid/air meniscuses in equilibrium with saturated vapour would necessarily be plane, immersion in the liquid could cause no further effect than saturation with the vapour, and we should have $x_L = x_s$.

If, as an alternative explanation, we suppose the expansion due to some form of chemical bonding, then again x_s and x_L should be equal, this time on account of the necessary equality of the chemical potentials in the liquid and saturated vapour phases.

The implications carried by the relation $x = \lambda F$ and the theory of the expansion effect previously given are in general different. The surface energy decrement F_s caused by the contact of a plane solid with a saturated vapour, and F_L, that caused by immersion in the liquid, are related by the equation

$$F_L = F_s + \gamma_{LV} \cos\theta, \tag{7}$$

where γ_{LV} is the surface energy of the liquid and θ the equilibrium angle of contact (wetting angle).

The relation to be expected theoretically between x_L and x_s on the basis of equation (1) depends upon the model adopted to represent the structure of the charcoal, and also on the magnitude of θ for the liquid considered. If the charcoal is conceived as a network of fine capillaries with very small external openings, and if the walls of these are perfectly wettable ($\theta = 0$), then capillary condensation would necessarily occur, and we should have the state of affairs already considered above ($x_L = x_s$). With a finite θ, on the other hand, the saturated adsorbed films and the normal liquid would necessarily constitute separate and distinct phases (Frumkin and others 1932); capillary condensation (as ordinarily understood) might then be precluded by the inability of the adsorbed films to act as nuclei for liquid condensation. Experiment has shown that wood charcoal can be exposed to *supersaturated* methyl alcohol vapour without its macropores becoming filled with liquid, and a similar negative result was obtained when a charcoal block was suspended in the fully saturated vapour of this substance for a long period (Bangham and Razouk 1937). Contact with *liquid* methyl alcohol under these circumstances caused a gradual filling of the pores against gravity, this process taking place at a rate indicating that with this liquid $0° < \theta < 90°$.*

Even in the absence of capillary condensation, equal values of x_L and x_s could result if the immersing liquid were unable to penetrate the finer capillaries forming the major part of the adsorbing surface. Relationship (iii), on the other hand, could probably be reconciled only with a much more open structure, for it is hard to believe that films adsorbed on mercury could behave similarly to an adsorbed phase more or less imprisoned in irregular capillaries of little more than molecular dimensions.

It will be observed that even supposing the liquid to have full access to the adsorbing surface, equations (1) and (7) set an upper limit to the expansion difference $(x_L - x_s)$ when the charcoal, already saturated, is plunged into the liquid. Writing $x_L = \lambda F_L$, $x_s = \lambda F_s$, equation (7) gives us

$$x_L - x_s = \lambda(F_L - F_s) = \lambda \gamma_{LV} \cos\theta. \tag{8}$$

Since for methyl alcohol $0° < \theta < 90°$, whilst γ_{LV} (25° C.) is *ca.* 22 ergs/cm.^{2}, it follows that $\gamma_{LV} \cos\theta < 22$ ergs/cm.^{2}. This is but a small fraction of F_s, the value of which, calculated on the basis of relationship (iii) above, is rather more than 200 ergs/cm.^{2}. *An observed difference $(x_L - x_s)$ exceeding, say, one-tenth of the saturation expansion would thus be incompatible with the calculations of specific surface, film thickness, and surface energy lowering previously given.* The difference found experimentally varies between one-sixteenth and one-twelfth of the saturation values, according to the temperature.

Heat of wetting. The above discussion makes it clear that it is not F_s but F_L that must be introduced into the Gibbs-Helmholtz equation to obtain a correct expression for the heat of *wetting*. If 1 g. of hitherto uncontaminated material is immersed in a liquid the heat evolved per cm.^{2} is $\left(F_L - \frac{T dF_L}{dT}\right)$;

---

* The slow rate of capillary rise would suggest a larger contact angle for methyl alcohol than is consistent with the marked difference between x_L and x_s found experimentally. The structure of charcoal is such, however, that measurements of the rate of capillary rise can be of no more than qualitative significance in this connexion.

and if the wetting extends all over the adsorbing surface of Σ cm.^{2}, the heat of wetting $(-\Delta H)$ is given by

$$(-\Delta H) = \Sigma\left(F_L - \frac{T dF_L}{dT}\right). \tag{9}$$

The assumption that $x_L = \lambda F_L$ and that λ is invariant with temperature leads to the equation

$$(-\Delta H) = \frac{\Sigma}{\lambda}\left(x_L - \frac{T dx_L}{dT}\right), \tag{10}$$

which lends itself to direct experimental test.

Scope of investigation. The experiments to be described aimed at finding, for a given experimental liquid, (1) the value of the ratio λ/Σ from expansion-adsorption-pressure data; (2) the saturation and immersion expansions x_s and x_L, each at a series of temperatures; and (3) the heat of immersional wetting. It will be readily understood that in order to provide a valid test of (10), and of the constancy or otherwise of λ, these data had necessarily to refer all to the same charcoal, and in the same physical condition.

APPARATUS AND MATERIALS

The choice of an experimental liquid was limited by the requirement that the full heat of wetting should be liberated within a minute or so of the immersion of the charcoal, i.e. before the corrections for radiation and conduction became serious. Even with highly evacuated charcoals, sealed in tubes in the manner to be described, only two liquids, methyl alcohol and water, were found to satisfy this requirement. The heat of wetting with water being too small to be measured with the required accuracy, methyl alcohol was chosen.

The charcoal and extensometer. Since it was desired to immerse the extensometer entirely in the liquid, as well as to carry out other operations liable to cause a slight amount of mechanical shock, a rather dense charcoal was used, and mounted in a highly robust form of extensometer. Highly porous charcoals, showing a larger swelling effect, are extremely fragile and somewhat liable to disintegrate when expanded.

A large compact-looking piece of German willow charcoal (charcoal IV; block density 0·366 g./c.c.) was exhausted for some days at 560° C. in an all-glass apparatus which comprised a mercury vapour pump and McLeod gauge. After cooling, a rod *ca.* 9 × 9 × 60 mm. was cut from it for use in the extensometer, the remainder serving for heat of wetting determinations.

The extensometer, a nickel-coated brass instrument designed and made by Mr J. Fliege, is sketched in fig. 1. Special attention was paid, in its construction, to eliminating the sources of error which may have affected the largest expansion values found with earlier models of more limited working range. The charcoal is clamped between two recessed cylinders *A* and *B* by strong spiral springs of nichrome. The upper cylinder *A* is locked to the body of the instrument by a screw. The lower one, a sliding fit in the body (its rotation is prevented by a rib fitting into a longitudinal slot in the cylinder), is supported only by the pressure of the springs. Its movement is transmitted to the long pointer *CD* by the upper of the two steel knife-edges at *D*, against which the base-plate of the pointer is pressed by the nichrome spring at *E*. The lower knife-edge is fixed.

Important points of design are the use of very strong springs which act directly on the cylinders enclosing the charcoal, and not on the knife edges; the long pointer, which enables the knife-edges to be set reasonably far apart even for a large magnification ratio; and the very strong setting of the knife-edges, which are acted upon only by relatively light stresses.

The extensometer is supported in its containing tube by the leaf springs at *F* and *G*, which press against the walls of the containing vessel.

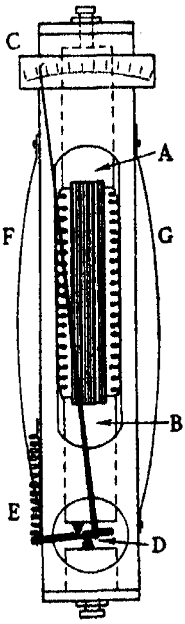


FIG. 1

The instrument was calibrated in two ways. A brass rod being substituted for the charcoal, the height of the upper cylinder was varied by turning the screw in the thread; the vertical travel of the upper knife-edge was then measured with an eyepiece micrometer, and compared with the change of scale reading of the pointer. In the second method the top plate of the instrument was removed, and a collar carrying an accurately threaded small-pitched screw with a large graduated head was substituted for the upper cylinder. These two calibrations gave results in excellent agreement, and showed the extensometer, when set for the experiment, to have a magnification ratio of 55; this was uniform over the whole of the expected range of movement, and it remained constant, as also did the zero-point of the instrument, even when the latter was submitted to a certain amount of rough handling.

Pointer observations were made with a vernier microscope reading to 0·01 mm. Variations of the length of the charcoal of the order of 0·0005 % were thus observable.

Change of zero-point with temperature: the linear expansion coefficient of gas-free charcoal. A disadvantage of the extensometer was the rather large change of zero-point with temperature, due to the very different expansion coefficients of brass and charcoal. Attempts to reduce this by various compensating devices having proved successful only at some sacrifice of accuracy, it remained to make an accurate determination of the temperature correction to be applied.

The extensometer and charcoal were first thoroughly degassed at 300° C. and the temperature was then allowed to fall by stages to 0° C. At each stage a constant temperature was maintained for 48 hr., the pointer reading having then been constant for some time. The scale readings plotted against temperature gave a good linear graph from which the length of the vapour-free charcoal at each temperature could be read off with accuracy.

Over the range 120–0° C. the difference between the coefficients of linear expansion of brass and charcoal was found to be $14{\cdot}4 \times 10^{-6}$. If the coefficient for brass is taken as $18{\cdot}9 \times 10^{-6}$, the value found for charcoal is $4{\cdot}5 \times 10^{-6}$, as against $5{\cdot}4 \times 10^{-6}$ recorded in the literature for "gas carbon". We believe this to be the first recorded expansion coefficient for a charcoal under conditions of high vacuum.

Expansion-adsorption-pressure measurements. The apparatus used was similar to that described in Part II. Full data were obtained for three isothermals, at 30, 17·5 and 0° C. In the experiments at the two lower temperatures each addition of vapour from the burette was made with the system warmed to 40° C., the extensometer tube being subsequently cooled down to the required temperature, and allowed to stand overnight before taking readings. This procedure was adopted in order that the pointer of the extensometer, notwithstanding the retrograde effect observable with methyl alcohol at lower temperatures, should approach each equilibrium position on a forward excursion.

After each experiment the charcoal was exhausted at 300° C.; on cooling the extensometer returned to its original zero-point.

Saturation and immersion expansions. The bathing of a solid in an exactly saturated atmosphere is commonly supposed to present a difficult practical problem, and the question as to whether an "elastic" gel (lyogel) reaches the same state of swelling in a saturated vapour as it does in the corresponding liquid is one on which opinion is still divided (see, for example, Katz 1933; Sheppard and Newsome 1934).

Though previous work had led us to anticipate little difficulty in getting accurate and reproducible values for the saturation expansion of charcoal, it was decided to render the control of temperature as complete as possible by detaching the adsorption vessel and immersing it *completely* in a thermostat. This procedure increased the range of temperatures over which measurements could be made.

Great precaution was taken to avoid accidental distillation of liquid on to the extensometer system. As far as concerns the charcoal itself this was probably unnecessary, for a similar charcoal has been exposed to supersaturated methyl alcohol vapour without causing condensation of liquid (Bangham and Razouk 1937). The danger lay in the condensation of drops on the readily wettable metal surfaces of the extensometer, and their subsequent movement into the interstices of the charcoal once contact had been made. The operation was carried out as follows.

The tube connecting the extensometer vessel to the vapour burette was first drawn to a thick-walled capillary neck, and a small appendix fused on just below this. When the charcoal had been given the usual vacuum heat treatment, it was exposed to the vapour of the methyl alcohol, and an excess of this was distilled into the appendix. The appendix being kept in ice-water in a small Dewar vessel, the whole extensometer tube was now sealed off at the capillary neck, and immersed completely, with the Dewar vessel, in a thermostat at 45° C. The following day, when the extensometer had come to temperature equilibrium with the thermostat, the Dewar vessel was removed. The system was then left for several days to make sure that no further movement of the extensometer would take place. The temperature of the thermostat tank was then reduced in stages to 0° C., at least 48 hr. being allowed the system to come to equilibrium (24 hr. more than sufficed for this). This procedure ensured that the extensometer should always approach its equilibrium on a forward excursion.

The saturation expansions so measured were found to be accurately reproducible in different experiments (as, for example, after the immersion expansions had been measured), and there can be no doubt that they represent the final equilibrium reached with the saturated vapour.

For the immersion experiments a large supply of methyl alcohol was purified and distilled *in vacuo* until quite free from dissolved gases. The extensometer tube, after the usual evacuation, was first exposed to the vapour, and then filled with liquid until the charcoal was completely covered. The tube was then sealed off and immersed in a thermostat, the temperature of which was lowered in stages, just as in the experiment with the saturated vapour.

Measurement of heat of wetting. A full account of the technique adopted will be given elsewhere; it followed in a general way that of Patrick and Grim

(1921). An important innovation was the enclosure of the charcoal, evacuated under the usual standard conditions, in a sealed tube of which the end could be broken under the wetting liquid. A small correction was applied for the dissipation, as heat, of the kinetic energy of the liquid thus forced into the tube.

In each experiment the thermal capacity of the system was measured by passing an electric current through an accurately standardized heater coil immersed in the liquid, and determining the e.m.f. potentiometrically.

RESULTS AND DISCUSSION

The ratio λ/Σ *for charcoal* IV. Except that the constants λ and Σ were different from those of the charcoal previously used, and that no attempt was made to get data for the unstable ("primary") films formed by methyl alcohol, the expansion-adsorption-pressure measurements substantially confirm the results given in Part III. The data need not, therefore, be recorded in detail. Assuming the expansion directly proportional to the surface energy lowering, the Gibbs equation is accurately valid over four-fifths of the range of expansion values; before this stage is reached there is evidence of a change of phase, associated with a discontinuity.

With the aid of equation (3), values of λ/Σ were calculated at 30, 17·5, and 0° C.; this could be done very accurately by evaluating $\int s\,d\log_e p$ graphically over the range of validity of the equation, and comparing the result with the corresponding change in x. The results are as follows:

$t = 30°$ C.	$RT\lambda/\Sigma = 37{\cdot}7$ %	$\lambda/\Sigma = 1{\cdot}50 \times 10^{-9}$ ergs^{-1}
17·5	35·2	1·46
0	33·6	1·48

Recorded in the second column are limiting values of the percentage expansion per mole ("molecular expansion") as the adsorption approaches zero. These are roughly proportional to the absolute temperature; those of the ratio λ/Σ given in the last column may clearly be taken as temperature-invariant.

Fig. 2 shows a part of the graph obtained by plotting the expansion (abscissae) against $\log_{10} p/p_0$ (p_0 = saturation pressure) for two separate experiments carried out at 30° C. The values of the saturation and immersion expansions, determined in separate experiments after the adsorption vessel had been sealed off and completely immersed in a thermostat at 30° C., are also inserted in the figure (double circles). It will be observed that whilst the saturation value x_s lies reasonably near the curve resulting from the expansion-adsorption-pressure experiments, the liquid immersion value x_L on the contrary does not do so. To bring the latter into line with the data for unsaturated vapours it would be necessary to assume the relative pressures subject to an error of no less than 12 %, in spite of the complete agreement between the observed saturation pressures and those recorded in the standard tables. The observed differences between x_L and x_s thus exceed by far any discrepancies due to experimental error.

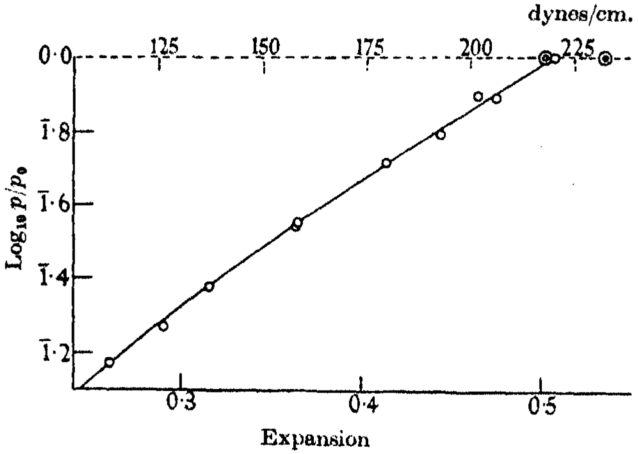


FIG. 2. Graph of $\log_{10} p/p_0$ against expansion (abscissae) at 30° C. The double circle on the left is the saturation expansion (mean value) at this temperature; that on the right is the expansion on immersion. At the top is a scale of dynes/cm. of surface tension lowering.

The saturation and immersion expansions at different temperatures are shown in fig. 3. The graph of x_s is linear over the range of temperatures examined; if produced to cut the abscissa axis it would do so above 500° C., i.e. far above the critical temperature of the alcohol.

The x_L graph on the other hand is convex to the temperature axis, and, though experiments over a wider range of temperatures are clearly desirable, the departure from linearity certainly exceeds any possible error of measurement. Within the experimental range the vertical intercept between the immersion and saturation curves actually passes through a minimum.

Calculation of heats of saturation and of wetting. With the aid of the Gibbs-Helmholtz equations and the values of λ/Σ given above, the data of fig. 3 can be made to yield values for the heat of saturation and the heat of wetting.

The approximate linearity of the x_s graph means that the entropy of saturation,* given by

$$-\Sigma\frac{dF_s}{dT} = -\frac{\Sigma dx_s}{\lambda dT}$$

is nearly independent of the temperature, and so also the total energy† of saturation. These relations are not likely to hold over an extended range of temperature. The energy of saturation calculated amounts to 12·9 cal./g.

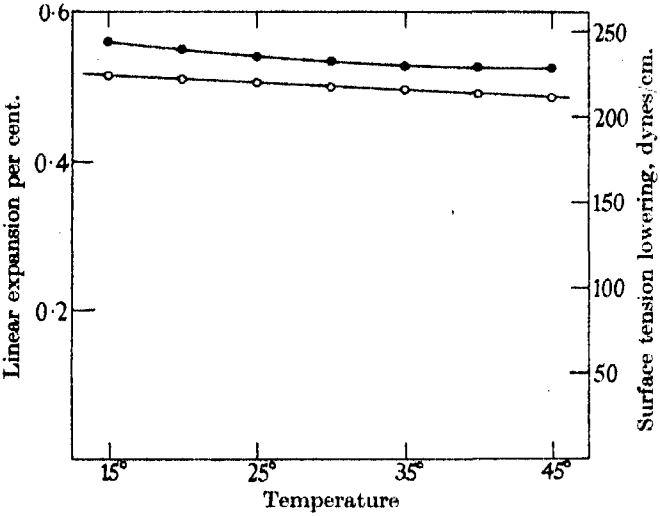


FIG. 3. Saturation expansions (white circles) and immersional expansions (black circles) at different temperatures. On the right of the diagram is a scale of dynes/cm. of surface tension lowering.

The entropy and energy of wetting,‡ on the other hand, should decrease with rise of temperature, since $-dx_L/dT$ decreases. To obtain a valid estimate of the heat of wetting this coefficient must be accurately evaluated, and this is best done in the middle of the temperature range, i.e at about 30° C. Two methods of treating the data, one algebraic and the other graphical, gave results in agreement, the value found for dx_L/dT at this temperature being $-0{\cdot}00140$ % per degree. Using now the Gibbs-Helmholtz relation in the form given by equation (10), we find 15·8 cal./g. as the calculated heat of wetting at 30° C.

---

* Increase of entropy of a system comprising 1 g. of evacuated charcoal and some *liquid* methyl alcohol, when the vapour of the latter is allowed to saturate the charcoal; it will be recalled that F_s is not the surface free energy, but its decrement due to adsorption.

† *Decrease* of energy of system under these conditions.

‡ Similarly defined to entropy and energy of saturation.

Measured heat of wetting. Three calorimetric measurements of the heat of wetting of charcoal IV by methyl alcohol at 30° C., using the vacuum technique described, gave 16·3, 15·7 and 15·9 cal./g. The mean value of 16·0 cal./g. is in excellent agreement with that calculated from the extensometric data.

Similar measurements at 20 and 40° C. show that the heat of wetting decreases appreciably with rise of temperature. The mean values found were:

$t = 20°$	$30°$	$40°$ C.
$(-\Delta H) = 16{\cdot}6$	$16{\cdot}0$	$15{\cdot}7$ cal./g.

Since a decrease of $(-\Delta H)$ with temperature necessarily means that d^2F_L/dT^2 is positive (this is readily seen on differentiating the Gibbs-Helmholtz equation with respect to T), the curve of F_L against T must be convex to the temperature axis. This again confirms the evidence of the extensometric data of fig. 3.

Structure and specific surface of charcoal IV. In the introduction to this paper suspicion was cast on relationship (iii) (which refers to the similarity of behaviour of films on charcoal and on mercury), on account of the very reasonable doubt as to whether the films could be regarded as bounded only on one side by the adsorbing surface of the charcoal. This suspicion is substantially removed by the evidence, now brought forward, that on immersion in methyl alcohol the charcoal behaves as if the liquid has access to the whole adsorbing surface. The observed differences between the immersion and saturation expansions lie within the narrow limits set by equation (8); the corresponding surface energy difference is greater than would be possible, according to equation (7), if no more than a small fraction of the surface became wetted. The results obtained—and the absence of capillary condensation—are most readily explained on the basis of a micellar structure for charcoal. It does not follow, of course, that with more active and finer structured charcoals the effects due to surface irregularities could be equally well disregarded.

Accepting relationship (iii), it becomes possible to evaluate λ and Σ and to assign values to the free and total energy changes per unit surface which accompany saturation and wetting.

Calculations along the lines indicated in Part IV, using, in conjunction with our expansion-adsorption data for methyl alcohol, the values given by Cassel and Salditt (1931) for the adsorption of this substance on mercury at $F = 30$ and $F = 35$ ergs/cm.^{2} lead respectively to $\Sigma = 149$ m.^{2} and $\Sigma = 135$ m.^{2} for charcoal IV. These estimates depend so much, however, on the accuracy of individual observations that they must be treated with reserve; moreover, it is by no means certain whether the transition of the film to the form stable at higher concentrations is complete at this stage, and this renders adsorption values calculated with the aid of Gibbs's equation untrustworthy. A better plan is to compare the adsorbing capacities, under comparable conditions, of charcoals IV and II; the specific surface assigned to the latter (*vide* Parts III and IV) is the mean of several concordant determinations, and rests on the general coincidence of the mercury and the charcoal data rather than on individual observations. This indirect method of calculation leads to $\Sigma = 155$ m.²; the other values given above approximate as closely to this as can be expected under the circumstances.

Saturation and immersional energy changes per unit area. Taking $\Sigma = 155$ m.² and $\lambda/\Sigma = 1{\cdot}48 \times 10^{-9}$ ergs^{-1}, the value found for λ is 0·00229 % per dyne/cm. On the basis of this estimate, a scale of dynes has been inserted in fig. 2 (at the top), and in fig. 3 (right-hand side).

Referred to unit area of adsorbing surface, the surface energy decrements accompanying saturation and immersion of charcoal in methyl alcohol, and the total energy of wetting are as follows:

$t = 20°$ C.	$F_s = 223$ ergs/cm.²	$F_L = 240$ ergs/cm.²	$(-\Delta H)/\Sigma = 449$ ergs/cm.²
30	218	234	433
40	214	230	425

The F_s values (second column) are a little higher than those arrived at from earlier data, but agree well with that obtained by integrating Coolidge's 59·5° C. isotherm for methyl alcohol (Bangham and Razouk 1937); this is of interest because Coolidge (1924) used a steam-activated charcoal. The total energies of wetting (last column) were obtained by dividing the calorimetrically determined heats by the estimated value of Σ.

It is our opinion that the energy changes recorded above represent real properties of the charcoal-methyl alcohol system, and are not greatly affected by impurities which may have been present in the charcoal. To influence fundamentally an adsorbing surface of the order of 155 m.², an impurity would have to be present in very considerable quantity. Analysis of similar charcoals after strong vacuum-heating has shown a small proportion of hydrogen to be present, and a very little oxygen; but the influence of these on so great a surface could be only very limited. The quantity of ash in these charcoals is extremely small.

With charcoals which have been submitted to some of the various activation processes it would become impossible, no doubt, to associate a definite energy change with a defined area of adsorbing surface; but this state of affairs is in no way foreshadowed in the results up to now presented. With the inactive charcoal used the energy relations characteristic of plane surfaces appear to be valid within experimental error.

The senior author gratefully acknowledges the award of a Leverhulme Research Fellowship, which has materially expedited the publication of this work. The experiments described were carried out in the laboratories of the Department of Chemistry, Egyptian University, Cairo.

SUMMARY

The percentage linear expansion x_s of a rod of inactive wood charcoal when exposed to the exactly saturated vapour of methyl alcohol is found to be less than the expansion x_L produced by immersion in the liquid. The expansion is proportional to the surface energy lowering of the charcoal, and this relation is confirmed by comparing $\left(x_L - T\frac{dx_L}{dT}\right)$ with the measured heat of wetting. On immersion of the air-free charcoal the liquid wets all, or nearly all, of the adsorbing surface. From well-supported estimates, previously given, of the specific surface of the charcoal, it becomes possible to assign values to the free and total energy changes per cm.^{2} which accompany saturation and immersion.

REFERENCES

Bangham 1934 *Proc. Roy. Soc.* A, **147**, 175.
— 1937 *Trans. Faraday Soc.* **33**, 805.
Bangham, Fakhoury and Mohamed 1932 *Proc. Roy. Soc.* A, **138**, 162.
— — — 1934 *Proc. Roy. Soc.* A, **147**, 152.
Bangham and Razouk 1937 *Trans. Faraday Soc.* **33**, 1459, 1463.
Bartell and Fu 1930 *Colloid Symp. Annu.* **7**, 135.
Cassel and Salditt 1931 *Z. phys. Chem.* **155**, 321.
Coolidge 1924 *J. Amer. Chem. Soc.* **46**, 596.
Freysinnet 1936 "Une Révolution dans les Techniques du Béton." Paris.
Frumkin, Gorodetzkaja, Kabanow and Nekrassow 1932 *Phys. Z. Sowjet.* **1**, 255.
Harkins and Ewing 1921 *J. Amer. Chem. Soc.* **43**, 1787.
Katz 1933 *Trans. Faraday Soc.* **29**, 279.
Patrick and Grim 1921 *J. Amer. Chem. Soc.* **43**, 2144.
Sheppard and Newsome 1934 *Industr. Engng Chem.* **26**, 285.
Whipple 1922 "Physical Society General Discussion on Hygrometry." London.
Williams 1918 *Proc. Roy. Soc. Edinb.* **38**, 23.

INDEX TO VOLUME CLXVI (A)

Adsorption, heat of (Bangham and Mosallam), 558.
Allibone (T. E.) and Meek (J. M.) The development of the spark discharge, 97.
Arnot (F. L.) and M'Ewen (Marjorie B.) The formation of helium molecules, 543.

Bangham (D. H.) and Mosallam (S.) The adsorption of vapours at plane surfaces of mica. II. Heats of adsorption and the structure of multimolecular films, 558.
Bangham (D. H.) and Razouk (R. I.) The swelling of charcoal. Part V. The saturation and immersion expansions and the heat of wetting, 572.
Bhabha (H. J.) On the theory of heavy electrons and nuclear forces, 501.
Blackman (M.) On the diamagnetic susceptibility of bismuth, 1.
Born (Max) Application of "reciprocity" to nuclei, 552.
Born (Max) and Fuchs (Klaus) The statistical mechanics of condensing systems, 391.
Bradley (A. J.) and Taylor (A.) An X-ray study of the iron-nickel-aluminium ternary equilibrium diagram, 353.
β-Transitions in a coulomb field (Hoyle), 249.

Chang (W. Y.) *See* Smith and Chang.
Charcoal, swelling of (Bangham and Razouk), 572.
Clark (G. L.) *See* Eddington and Clark.
Cochrane (W.) Polish on metals, 228
Collens (H.) *See* Schonland, Hodges and Collens.
Condensing systems, statistical mechanics of (Born and Fuchs), 391.
Copper-gold alloy, atomic rearrangement (Jones and Sykes), 376.
Cosmic radiation, showers produced by (Heitler), 529.
Cosmic-ray particles, energy loss (Wilson), 482.
Coulson (N.) *See* Howell and Coulson.

de Graaff Hunter (J.) Some experimental observations for longitude made by theodolite fitted with a shutter eyepiece, 197.
Diamagnetic susceptibility of bismuth (Blackman), 1.
Diamagnetism of cadmium (Rao and Sriraman), 325.
Dielectric polarization of a *n*-long chain ketone (Müller), 316.
Dispersion formula for nuclear reactions (Kapur and Peierls), 277.

Eddington (Sir A.) and Clark (G. L.) The problem of *n* bodies in general relativity theory, 465.
Electronic showers, cascade theory (Landau and Rumer), 213.

Fröhlich (H.), Heitler (W.) and Kemmer (N.) On the nuclear forces and the magnetic moments of the neutron and the proton, 154.
Fuchs (Klaus) *See* Born and Fuchs.

Goodeve (C. F.) and Wood (L. J.) The photosensitivity of diphenylamine-*p*-diazonium sulphate by the method of photometric curves, 342.
Guggenheim's theory of binary liquid mixtures (Rushbrooke), 296.

Half-pitot tube, measurements with (Taylor), 476.
Hartree (D. R.) and Hartree (W.) Self-consistent field with exchange for potassium and argon, 450.
Hartree (W.) *See* Hartree and Hartree.
Havelock (T. H.) The lift and moment on a flat plate in a stream of finite width, 178.
Heavy electrons and nuclear forces, theory (Bhabha), 501.
Heitler (W.) Showers produced by the penetrating cosmic radiation, 529.
Heitler (W.) *See* Fröhlich, Heitler and Kemmer.
Helium molecules, formation of (Arnot and M'Ewen), 543.
Hodges (D. B.) *See* Schonland, Hodges and Collens.
Howell (H. G.) and Coulson (N.) The spectrum of thallium chloride, 238.
Hoyle (F.) β-Transitions in a coulomb field, 249.

Iron-nickel-aluminium ternary equilibrium diagram (Bradley and Taylor), 353.

Jones (F. W.) The measurement of particle size by the X-ray method, 16.
Jones (F. W.) and Sykes (C.) Atomic rearrangement process in the copper-gold alloy Cu_3Au. II., 376.

Kapur (P. L.) and Peierls (R.) The dispersion formula for nuclear reactions, 277.
Kemmer (N.) Quantum theory of Einstein-Bose particles and nuclear interaction, 127.
Kemmer (N.) *See* Fröhlich, Heitler and Kemmer.

Landau (L.) and Rumer (G.) The cascade theory of electronic showers, 213.
Lift and moment on a flat plate (Havelock), 178.
Light, theory of scattering (Mueller), 425.
Longitude observations by theodolite fitted with a shutter eyepiece (de Graaff Hunter), 197.
Lovell (A. C. B.) The electrical conductivity of thin metallic films. III. Alkali films with the properties of the normal metal, 270.

M'Ewen (Marjorie B.) *See* Arnot and M'Ewen.
Meek (J. M.) *See* Allibone and Meek.
Metallic films, electrical conductivity (Lovell), 270.
Misener (A. D.) The resistance of superconducting cylinders in a transverse magnetic field, 43.
Mosallam (S.) *See* Bangham and Mosallam.
Mueller (Hans) On the theory of scattering of light, 425.
Müller (Alex) The dielectric polarization of a *n*-long chain ketone at constant volume and variable temperature, 316.
Multimolecular films, structure of (Bangham and Mosallam), 558.

n-Bodies in general relativity theory (Eddington and Clark), 465.
Nuclear forces and magnetic moments of the neutron and the proton (Fröhlich, Heitler and Kemmer), 154.

Particle size measurement by the X-ray method (Jones), 16.
Peierls (R.) *See* Kapur and Peierls.
Photosensitivity of diphenylamine-*p*-diazonium (Goodeve and Wood), 342.
Polish on metals (Cochrane), 228.
Progressive lightning (Schonland, Hodges and Collens), 56.

Quantum theory of Einstein-Bose particles (Kemmer), 127.

Radioactive α-particles from ^{8}Li, range-distribution curve (Smith and Chang), 415.
Rao (S. Ramachandra) and Sriraman (S.) Diamagnetism of cadmium, 325.
Razouk (R. I.) *See* Bangham and Razouk.
"Reciprocity", application to nuclei (Born), 552.
Rumer (G.) *See* Landau and Rumer.
Rushbrooke (G. S.) A note on Guggenheim's theory of strictly regular binary liquid mixtures, 296.

Saturation and immersion expansions of charcoal (Bangham and Razouk), 572.
Schonland (B. F. J.), Hodges (D. B.) and Collens (H.) Progressive lightning. V. A comparison of photographic and electrical studies of the discharge process, 56.
Self-consistent field with exchange for potassium and argon (Hartree and Hartree), 450.
Shutter eyepiece theodolite, observations with (de Graaff Hunter), 197.
Smith (C. L.) and Chang (W. Y.) An accurate determination of the range-distribution curve of the radioactive α-particles from ^{8}Li, 415.
Spark discharge, development (Allibone and Meek), 97.
Sriraman (S.) *See* Rao and Sriraman.
Structure of animal fibres (Woods), 76.
Superconducting cylinders in a transverse magnetic field (Misener), 43.
Sykes (C.) *See* Jones and Sykes.

Taylor (A.) *See* Bradley and Taylor.
Taylor (G. I.) Measurements with a half-pitot tube, 476.
Thallium chloride, spectrum (Howell and Coulson), 238.

Vapours, adsorption of (Bangham and Mosallam), 558.

Wilson (J. G.) The energy loss of penetrating cosmic-ray particles in copper, 482.
Wood (L. J.) *See* Goodeve and Wood.
Woods (H. J.) X-ray studies of the structure of hair, wool and related fibres. IV. The molecular structure and elastic properties of the biological cells, 76.

X-ray studies of the structure of animal fibres (Woods), 76.

END OF THE ONE HUNDRED AND SIXTY-SIXTH VOLUME (SERIES A)

ABSTRACTS

OF PAPERS COMMUNICATED TO THE ROYAL SOCIETY OF LONDON

In accordance with a resolution of Council, summaries or abstracts of papers are to be published as soon as practicable. The publication of such abstracts in no way indicates that the papers have been accepted for publication in any fuller form. These abstracts will be issued for convenience with the "Proceedings of the Royal Society of London" but do not form a part of the "Proceedings".

---

19 MAY 1938

The investigation of electron lenses. By O. KLEMPERER and W. D. WRIGHT. (*Communicated by W. H. Eccles, F.R.S.—Received* 29 *March* 1938.)

The paper describes two methods of deriving the optical constants and spherical aberration of electrostatic lens systems. In the first, the electrical field is imagined to be divided into a series of equipotential surfaces at each of which there is a finite change of potential. To a first approximation these fields are spherical close to the axis and the normal optical ray tracing formulae can be used to compute the paths of electrons through the system. Results are reported for a two-tube lens for various focusing ratios, the two tubes being of equal diameter. The aberration contributions of the different refracting surfaces to the final aberration can be evaluated and the curvature of field can also be determined by this means. In the second method, the Hartmann test used in glass optics has been applied to the experimental investigation of electron lenses. Narrow pencils of electrons pass through a suitable diaphragm and their course when focused by the lens under test is found by measurements of their intersections with a fluorescent screen placed at appropriate points along the axis. The design of an electron gun suitable for these tests is described, and results for the two-tube lens are given for voltage ratios from 2·5:1 to 15:1.

Trypanocidal activity and chemical constitution. By G. M. DYSON, T. H. FAIRBROTHER, F. A. MASON and A. RENSHAW. (*Communicated by H. S. Raper, F.R.S.—Received* 29 *March* 1938.)

The physiological action of the complex ureas has been shown to be very specific. The smallest alteration in chemical constitution such as the omission of a methyl

group seriously alters the trypanocidal activity. The object of this research has been to investigate the variations in trypanocidal activity produced by certain changes in the typical urea structure; it has been shown that the accumulation of three sulphonic acid groups in suitable orientation in the terminal part of the molecule is necessary for the development of trypanocidal activity; further that the intermediate components must not depart appreciably from the *m*-amino-*p*-benzoyl-*m*-amino-*p*-toluyl structure. More latitude is allowable in the structure of the bridge linkage, which may have other configurations without complete extinction of trypanocidal activity. Whilst the urea bridge gives the best results, the carbazide linkage is almost equally suitable, although precautions must be taken not to destroy it with the formation of a urazine structure, during preparation.

Numerous examples of the variations in the typical structure of these ureas have been prepared and their physiological properties examined; they comprise the systematic alterations of terminal, intermediate and bridge linkages and also the introduction of halogen substituents in the aromatic rings. In addition, for purposes of comparison, a number of dyestuffs of the trypan red and blue series have been prepared and examined.

Experiments made with a half-Pitot tube. By G. I. TAYLOR, F.R.S.

(*Received* 31 *March* 1938.)

The half-Pitot tube was invented by Stanton for measuring the tangential stress at the surface of a solid past which fluid is flowing. Stanton calibrated his tubes by inserting them in a pipe through which air was flowing in steady motion. He found that the pressure in his tubes decreased as their size diminished, but he did not attain the limiting condition when the half-Pitot is very small, nor did he express his results in non-dimensional form so that observations made with different sized tubes could be compared.

The present paper describes measurements made with a half-Pitot tube in a stream of glycerine. It is found that when the size of the tube is very small the pressure in it is 1·2 times the tangential stress acting on the surface through which it projects. The results, together with those of Stanton, are expressed in non-dimensional form.

Summary of electro-magnetic induction in moving bodies. By E. G. CULLWICK.

(*Communicated by Sir Frank Smith, Sec.R.S.—Received* 24 *March* 1938.)

The paper has particular reference to the problems of unipolar induction.

The resultant electric field at a point is given in terms of: (*a*) the positions of electric charges, (*b*) the magnitude and velocity of each component of the magnetic field, on the assumption that each component is to be regarded as sharing the motion of its source, and (*c*) the accelerations of electric charges.

It is held that this method of calculation is consistent with the restricted theory of relativity, and removes the implication of a fixed aether from the usual relativity modifications of the Maxwell-Lorentz equations. It is further shown that open-circuit experiments on rotating magnets and solenoids are all consistent with the general validity of the moving-field theory, in spite of much argument to the contrary, and that indeed the balance of the evidence seems to be against the universal validity of the fixed-field Maxwell-Lorentz theory.

The hyperfine structure of the Zeeman components of the resonance lines of sodium.

By D. A. JACKSON and H. KUHN. (*Communicated by F. A. Lindemann, F.R.S.—Received* 4 *April* 1938.)

The hyperfine structure of the Zeeman π components of the two resonance lines of sodium was observed by means of the absorption in an atomic beam, the high resolving power instrument being a double etalon with separations of 2 and 8 cm. In strong fields, each Zeeman component was found to possess four hyperfine structure components, proving with certainty that the nuclear spin of sodium is $\frac{3}{2}$.

The alteration in the spacings of the lines as the field strength was changed from 3060 to 790 gauss was found to be in excellent agreement with that predicted by quantum mechanics.

A direct determination of the hyperfine structures of the terms $3S_{\frac{1}{2}}$ and $3^2P_{\frac{1}{2}}$ was made from the complete resolution of the line 5896.

Rotational analysis of the first negative band spectrum of oxygen.

By T. E. NEVIN. (*Communicated by A. W. Conway, F.R.S.—Received* 6 *April* 1938.)

The (1, 0), (0, 0) and (0, 1) bands of the first negative system of O_2^+ have been photographed in the second order of a 21 ft. grating. Each of the bands has been analysed into forty branches, the transition involved being $^4\Sigma \rightarrow {}^4\Pi$. The fine structure of the $^4\Sigma$ level agrees with the theoretical structure predicted by Budó. The $^4\Pi$ level is inverted. The structure of the level does not agree with that to be expected on theoretical grounds from formulae given by Brandt and Budó. It is suggested that this disagreement is caused by a perturbation similar to that observed by Bulthuis and Coster in the $A^2\Pi$ level of CO^+.

Studies on the nature of the amphibian organization centre. VII.

By C. H. WADDINGTON. (*Communicated by Sir Frederick Hopkins, F.R.S.—Received* 15 *January* 1938.)

1. Certain carcinogenic and oestrogenic substances, and other related compounds, were tested for evocator capacity by implantation in the blastocoel of newt gastrulae.

2. The substances were implanted in the form of suspensions in coagulated egg albumen. In an extensive series of controls, no inductions were obtained with albumen alone.

3. In the experimental series, a large number of inductions were obtained, the eggs showing much greater reactivity than in previous years. The results are summarized in Table I.

4. Inductions were obtained with a considerable number of substances, and no hypothesis can yet be made as to the relation between chemical structure and evocator power.

The origin and developmental mechanics of the avian sternum. By Honor B. Fell.

(*Communicated by E. S. Goodrich, F.R.S.—Received* 21 *April* 1938.)

The development of the avian sternum was studied in relation to three main problems: (1) the origin of the sternum, (2) its anatomical differentiation and (3) the union of the sternal plates. The normal development of the budgerigar sternum was investigated, and points which could not be settled by ordinary morphological methods were studied experimentally by tissue culture. It was shown that the presumptive sternal tissue is not derived from either the ribs or the coracoid although it develops in close association with the latter; its position in the lateral body wall of the 4-day embryonic fowl was defined experimentally. The anatomical development of the sternum was found to be controlled largely by intrinsic factors, but the results suggested that extrinsic factors probably serve (*a*) to provide the optimal conditions for the expression of potencies inherent in the rudiment, and (*b*) to maintain the normal anatomical structure after it has once differentiated. The shape of the sternum depends partly on the relationship of growth rate to differentiation rate, and artificial disturbance of this relationship produces abnormalities of form. Evidence was obtained indicating that union of the sternal plates is due primarily to the active, amoeboid migration of adjacent undifferentiated cells towards the mid-line, and secondarily to shrinkage and degeneration in the mid-ventral body wall.

Investigation into the relation of shower frequency to general cosmic-ray activity. By L. Jánossy.

(*Communicated by P. M. S. Blackett, F.R.S. —Received* 12 *April* 1938.)

The frequency of cosmic-ray showers was observed at sea level and under 30 m. of clay (60 m. of water equivalent). The shower frequencies at the two levels were compared with the corresponding vertical intensities. Part of the experiments were carried out with layers of light material close above the lead in order to maintain similar geometrical conditions for the shower production at both levels. The marked observed effect of the stone on the transition curve can be understood as due mainly to the spatial association of the shower-producing electrons emerging from the stone rather than to filtration. The relative shower intensity, i.e. (shower frequency/vertical intensity), of the showers producing the first slope of the transition curve was found to be mainly the same at sea level and underground; on the other hand the shower frequency due to showers producing the tail of the transition curve increases relatively with depth. The conclusion was drawn that most of the showers at sea level and underground are due to secondaries of some kind produced by the penetrating particles, and that consequently relatively few can be due to primary electrons.

ABSTRACTS

OF PAPERS COMMUNICATED TO THE ROYAL SOCIETY OF LONDON

In accordance with a resolution of Council, summaries or abstracts of papers are to be published as soon as practicable. The publication of such abstracts in no way indicates that the papers have been accepted for publication in any fuller form. These abstracts will be issued for convenience with the "Proceedings of the Royal Society of London" but do not form a part of the "Proceedings".

---

3 JUNE 1938

The absorption spectra of the halogen acids in the vacuum ultra-violet.

By W. C. PRICE. (*Communicated by R. G. W. Norrish, F.R.S.—Received* 20 *April* 1938.)

The absorption spectra of hydrogen iodide, bromide and chloride have been investigated in the vacuum ultra-violet. The bands observed are due to the excitation of a non-bonding $p\pi$ electron and are the strong resonance bands approaching the $^2\Pi_{1\frac{1}{2},\frac{1}{2}}$ states of the molecular ion. They start at 1762, 1491 and 1331 A for hydrogen iodide, bromide and chloride respectively. In the case of hydrogen iodide a well-developed Rydberg series of bands was found approaching the $^2\Pi_{\frac{1}{2}}$ state of HI^+, the limit of which corresponded to $10{\cdot}990 \pm 0{\cdot}005$ V. It was not possible to establish with certainty Rydberg series going to $^2\Pi_{1\frac{1}{2}}$ of HI^+ or to any of the $^2\Pi$ states of HBr^+ or HCl^+. However, from correlations made amongst the bands of the different acids and with the corresponding transitions in the methyl halides, the mean ionization potentials (i.e. $HX:{}^1\Sigma \rightarrow HX^+:{}^2\Pi$) are determined to be 10·71, 12·04 and 12·84 V for hydrogen iodide, bromide and chloride respectively, the probable error being about a tenth of a volt.

The influence of magnetic fields on persistent currents in superconductors.

By K. C. MANN, H. GRAYSON SMITH and J. O. WILHELM. (*Communicated by A. S. Eve, F.R.S.—Received* 20 *April* 1938.)

Using the method of persistent currents to detect the onset of resistance of cylindrical specimens in transverse and longitudinal fields, the critical fields H_T and H_L (marking the appearance of this resistance) have been investigated. H_L for spectroscopically pure tin agrees very well with published critical fields, but is slightly higher than these at lower temperatures. For spectroscopically pure tin, $\frac{H_T}{H_L}$ is a constant 0·58 independent of temperature. Supplementary measurements carried out on a single crystal of "chemically pure" tin give longitudinal critical fields H_L considerably lower than for the spectroscopically pure tin, and the characteristic ratio shows a linear dependence on temperature. It is concluded from these results and from the results of other workers, that the dependence of the ratio on temperature is due to some other factor than crystal structure or purity.

The critical fields for very pure tin to which was added small amounts of both supraconducting and non-supraconducting impurities were also investigated. In all cases, H_L for these impure specimens is lower than the value for pure tin, but the magnitude of the effect depends upon the type of impurity used. The effect of impurities on H_T is very marked. Supraconductors raise this critical field, while the non-supraconductor lowers it. The characteristic ratio for tin-bismuth is about the same as for pure tin, while for tin-lead and tin-thallium, the ratios are roughly 0·75 and 0·65 respectively.

The experiments definitely show that pure tin in the intermediate state can carry without resistance a considerable, though limited current. It is concluded that impurities affect this current-carrying capacity.

Time effects were also observed with all specimens as they passed into the intermediate state. These were caused by the dying away of persistent currents which had been induced in the circuit above the critical value that the circuit could carry without showing resistance. These effects also differed with the purity of the specimen, showing that the resistance of the intermediate state is affected by impurities.

A new form of resorcinol. Part I. Structure determination by X-rays.

By J. MONTEATH ROBERTSON and A. R. UBBELOHDE. (*Communicated by Sir William Bragg, P.R.S.—Received* 23 *April* 1938.)

The structure of the dense, high-temperature modification of resorcinol has been fully determined by means of a quantitative X-ray investigation. The structure is characterized by a more compact and parallel arrangement of the molecules than in ordinary resorcinol, which is a rather open arrangement. The benzene rings are apparently regular hexagons to within 0·05 A, but the OH groups are definitely displaced from the symmetrical positions, and one of them at least is also displaced out of the plane on the ring. The carbon-oxygen distances are 1·36–1·37 A, as in ordinary resorcinol. The orientation of the molecule in the crystal and the atomic co-ordinates are given. Hydroxyl bonds of length 2·70 and 2·75 A connect neighbouring molecules in the crystal, but the angles between these bonds are further distorted from the tetrahedral value than in ordinary resorcinol. All other intermolecular distances are greater than 3·4 A.

A new form of resorcinol. Part II. Thermodynamic properties in relation to structure. By J. Monteath Robertson and A. R. Ubbelohde. (*Communicated by Sir William Bragg, P.R.S.—Received* 23 *April* 1938.)

The anomalous temperature relationship between the polymorphs of resorcinol is discussed in the light of the crystal structures. The open α structure is maintained by hydroxyl bonds orienting the rigid resorcinol molecules. Transition to the denser β structure, which resembles that of many hydrocarbons, is associated with a distortion of the hydroxyl bonds, and a deformation of the resorcinol molecule which may involve a Mills Nixon effect.

Experiments are described on the preparation of α and β resorcinol, on the transition between the two forms, and on the heat of transition. The transition point of $C_6H_4(OH)_2$ is about 2° lower than that of $C_6H_4(OD)_2$, and the change from one form to the other is considerably more rapid in the deuterium compound. The heat of transition from α to β resorcinol is 220 ± 50 cal./mol. Experiments on the volume change from (supercooled) liquid resorcinol to either crystal structure show that there is only a small expansion on passing from the open α structure to the liquid. Experiments on the transition between two polymorphs of pentaerythritol show a normal temperature relationship for this substance.

The electrical response of the auditory mechanism in cold-blooded vertebrates. By E. D. Adrian, F.R.S., K. J. W. Craik and R. S. Sturdy. (*Received* 25 *April* 1938.)

The effects of sound on the ears of some cold-blooded vertebrates have been analysed by recording the electric changes set up in the inner ear and in the auditory nerve. The inner ear potentials (the equivalent of the Wever and Bray effect in the mammal) can be recorded in animals without a cochlea and seem to be due to the movement of polarized membranes. As they show the mechanical vibrations transmitted to the inner ear they can be used to compare the efficiency of different types of middle ear. Thus the alligator (with a membranous ear drum) gives an inner ear response to sounds of higher frequency than the tortoise (drum covered by external skin), though at low frequencies the tortoise ear is as sensitive. In fish the inner ear reacts to mechanical vibration although there is no middle ear.

A change of temperature has little effect on the performance of the middle ear or on the inner ear potentials, but the frequency at which the auditory nerve fibres can respond will be limited by the refractory period and therefore by the temperature. At 10° C. the nerve fibres in the frog and tortoise will not signal frequencies much above 200 per sec., whereas at 20° C. they will signal 400–500. These results emphasize the value of a constant high temperature for the receptor mechanism for sound.

The determination of the meteorological conditions of the atmosphere by the use of radio-sounding balloons.

By H. A. THOMAS. *(Communicated by Sir George Simpson, F.R.S.—Received* 26 *April* 1938.)

Existing methods of exploring the atmosphere by radio-sounding balloons are reviewed briefly and it is explained that arrangements involving variation of radio frequency are not likely to be satisfactory due to the wide frequency band required and the chances of appreciable interference. Furthermore, methods involving signalling codes are complex and expensive, and the delineation of the signals at the receiver presents considerable difficulty.

The essential features of an original method of exploration are described; in this method, a signal of fixed radio frequency is employed and each meteorological instrument produces a continuous variation of modulation frequency. Calibration of the meteorological parameters in terms of frequency can be made independently of the radio transmitter and measurement of these audio frequencies at the receiver is comparatively easy. The pressure and temperature measuring instruments are both arranged to produce variation of modulation frequency without the use of mechanical linkages. The cost of the apparatus is comparatively low and reproduction in large quantities is possible.

The results obtained from a number of experimental ascents are analysed and it is shown that a high degree of reliability and accuracy is obtainable. Using the present balloon, observation of both pressure and temperature up to an altitude of about 10 kilometres is obtained, the accuracy of the pressure and temperature determinations being about 5 mb. and 1° C. respectively.

The effect of anterior pituitary extracts on established lactation in the cow.

By S. J. FOLLEY and F. G. YOUNG. *(Communicated by A. S. Parkes, F.R.S.—Received* 26 *April* 1938.)

Single injections of a crude anterior pituitary extract, and of prolactin and thyrotrophic preparations obtained from fresh ox anterior lobe, into cows in declining lactation, resulted in each case in a substantial increase in the milk yield for 5 or 6 days. Only slight changes were found in the composition of the milk.

A single injection of prolactin prepared from desiccated ox anterior lobe administered in a dose equivalent, as regards crop gland stimulation, to the dose of prolactin from fresh gland, had no effect on the milk volume or composition. Likewise a single injection of a dose of the thyrotrophic fraction prepared from dried gland, equivalent to the dose previously used of the thyrotrophic fraction from fresh gland as regards ability to cause enlargement of the thyroid gland, had no effect on the milk volume or composition.

The daily injection of very large doses of prolactin prepared from dried gland resulted in a substantial and prolonged increase in milk volume, with a rise in milk lactose content accompanied by a fall in the milk chlorides.

The thyrotrophic preparation from fresh gland, although lactogenic in the lactating cow, possessed no detectable pigeon crop gland stimulating activity. The prolactin

preparation from dried gland, although possessing substantial pigeon crop gland stimulating activity, failed to exhibit lactogenic activity when a single dose was injected into a lactating cow. The relative lactogenic activities of the extracts used are paralleled, not by their activities in causing stimulation of the pigeon crop gland, but by their relative glycotropic (anti-insulin) activities.

There seems to be no satisfactory evidence that a single "lactation hormone" is secreted by the anterior pituitary gland and it is suggested that use of this term should be discontinued.

The mechanism of chain breaking in the decomposition of ethane. By C. N. HINSHELWOOD, F.R.S., and J. E. HOBBS. (*Received* 27 *April* 1938.)

In the decomposition of ethane there are chains which appear normally to be broken by the recombination of radicles, but, in the presence of nitric oxide, by reaction between it and the radicles. The equation for the curve showing the rate of reaction as a function of nitric oxide concentration contains constants which depend differently upon the ethane concentration according to the mechanism assumed for the normal recombination of the radicles. Experimental investigation gives evidence that the predominant mechanism for chain ending in the absence of nitric oxide is a ternary collision between two radicles and an ethane molecule, but that binary collisions are also in some degree effective. This result is discussed.

Reaction chains in the thermal decomposition of hydrocarbons. A comparison of methane, ethane, propane and hexane. By C. N. HINSHELWOOD, F.R.S., and J. E. HOBBS. (*Received* 27 *April* 1938.)

A comparison has been made of the inhibition by nitric oxide of the thermal decomposition of methane, ethane, propane and hexane. The mean chain length, as defined by the ratio of the rates of the uninhibited and fully inhibited reactions, shows no marked variation as the series is ascended, has values from 2 to 15 according to conditions, and, for a given hydrocarbon, decreases as the initial pressure increases. The curves showing the decrease in rate as a function of the nitric oxide concentration can be expressed in terms of a simple theory of chain breaking. The mean chain length depends upon the relative values of three activation energies, namely those of radicle formation, of decomposition by internal rearrangement, and of chain propagation. Approximate evaluation of the last shows that it tends to decrease with the higher hydrocarbons.

A study of sensitized explosions. II. Ignition phenomena in mixtures of carbon monoxide and oxygen sensitized by hydrogen. By E. J. Buckler and R. G. W. Norrish, F.R.S. (*Received* 27 *April* 1938.)

In extending the research of other workers on the catalytic influence of hydrogen on the oxidation of carbon monoxide, it has been found that ignition in mixtures containing hydrogen occurs at temperatures more than 150° below those of the pure gases, and is attended by complete combustion when the concentration of hydrogen exceeds about 0·5 %. The region of sensitized ignition is bounded by a lower pressure limit which was not investigated in detail, and an upper pressure limit the dependence of which on temperature and hydrogen concentration was determined by the slow withdrawal of sensitized mixtures from the reaction vessel up to the point of explosion.

Mixtures are stable above the upper pressure limit obtained by the withdrawal method, but may be ignited by sudden compression. A study of the minimum degree of compression required to produce inflammation in mixtures containing different inert gases proved that although the initial presence of free radicals containing hydrogen was essential, the primary cause of ignition was the momentary rise of temperature accompanying sudden compression.

It has also been found that mixtures containing a suitable percentage of hydrogen admitted suddenly to an evacuated reaction vessel ignite when the final pressure is only slightly greater than the upper limit determined by the withdrawal method, but for higher final pressures may be introduced without inflammation occurring. However, ignition reappears when the final pressure is further increased to values above 100 mm. Hg, due to approximately adiabatic compression of the first portions of gas to enter the reaction vessel. Owing to the momentary rise of temperature in the compressed portion of the mixture, ignition spreads out from reaction centres formed by chain branching processes operative during the early stages of admission.

A study of senzitized explosions. III. The kinetics of ignition of carbon monoxide and oxygen sensitized by hydrogen. By E. J. Buckler and R. G. W. Norrish, F.R.S. (*Received* 27 *April* 1938.)

It has been found that mixtures of carbon monoxide and oxygen containing only traces of hydrogen resemble oxy-hydrogen mixtures in that both ignite over the same range of temperatures, and both possess upper pressure limits of ignition not markedly affected by change of vessel diameter, which vary in the presence of inert gases in a manner suggesting that the processes responsible for multiplication of chain carriers may be prevented by collisions of the second kind with other molecules. In mixtures of $2CO + O_2$ of constant $CO : H_2$ ratio the relation between $1/T$ and $\log P$ (P being the pressure limit of ignition at $T°$ abs.) is linear when the concentration of hydrogen is less than about 2 %, with a slope corresponding to an apparent energy of activation of 13,300 cal., approximately one-half the corresponding figure given for mixtures of $2H_2 + O_2$, while for mixtures of $2CO + O_2$ containing higher percentages of hydrogen the curves show a variable slope. These facts suggest that carbon monoxide introduces addition chain-ending processes into the mechanism of the hydrogen-oxygen reaction. A similar value for the apparent energy of activation is obtained in mixtures of $2CO + O_2$ containing 0·5 % of deuterium, while the ratio of the values of the pressure limits of ignition of mixtures sensitized by the same percentage of hydrogen and deuterium respectively becomes constant at 1·42 when the concentration of sensitizer is below about 2 %.

The catalytic effect of hydrogen on the combustion of carbon monoxide, and the differences in kinetic characteristics between this reaction and that in mixtures of pure hydrogen and oxygen can be completely described by the introduction of two additional elementary reactions $O + CO + M = CO_2 + M$ and $OH + CO = CO_2 + H$ into the mechanism for the ignition of mixtures of $2H_2 + O_2$ proposed by Kassel and Storch (1935) and modified by von Elbe and Lewis (1937).

The origin of the erythema curve and the pharmacological action of ultra-violet radiation. By J. S. Mitchell, with an appendix by C. B. Allsopp and J. S. Mitchell. (*Communicated by E. K. Rideal, F.R.S.—Received* 2 *May* 1938.)

The measurements of the erythema curve and the relevant known facts of the reaction of the human skin to ultra-violet light are reviewed. The possibility of a photochemical mechanism is discussed and developed theoretically.

Quantitative evidence is presented to show that the erythemal response is due to photochemical decomposition of typical proteins in the stratum mucosum by the light transmitted through the stratum corneum; this latter behaves as an inert scattering and absorbing screen. The photochemical changes are limited to the stratum mucosum by the distribution of pigment. The latent period is, at least in part, the time required for the diffusion of active photo-decomposition products to the minute vessels.

The form of the erythema curve is calculated, within the limits of experimental error, from absorption measurements of a typical protein and of the human stratum corneum. The calculated mean value of the wave-length of maximum erythemal efficiency is 2975 (± 15) A, in close agreement with that observed. The known minimal erythema dose is consistent with the formation of photodecomposition products *in vitro* in concentration of a reasonable order of magnitude.

It is suggested that the H-colloid may be a proteose formed by oxidative photolysis of proteins.

The relationship of erythemogenesis to pigmentation and carcinogenesis by ultra-violet radiation is briefly discussed.

The function of the ampullae of Lorenzini, with some observations on the effect of temperature on sensory rhythms. By A. Sand. (*Communicated by S. W. Kemp, F.R.S.—Received* 2 *May* 1938.)

The histological structure of the ampullae of Lorenzini is still in dispute, though there is strong evidence that they are not neuromasts. The central connexions of their innervation have not been thoroughly studied. Their functional activity, as revealed by the present oscillographic investigation, is entirely distinct from that of lateral-line organs. For these reasons their unqualified inclusion in the acustico-lateralis system is unwarranted.

Like acustico-lateral receptors they possess an autonomous rhythm of activity, but in contrast with the former, they are unaffected by any form of mechanical stimulation. Their sensory sensitivity is confined to thermal stimulation.

Their responses to temperature changes are highly peculiar. They respond to cooling by an acceleration, to warming by an inhibition of the rhythm of impulse discharge. Such responses have been shown to occur throughout the physiological temperature range, from 3 to 20° approximately, though it is probable that they are most sensitive in the region 10–15°. In this region clear responses have been recorded to temperature differences of one or two-tenths of a degree.

After a response to a temperature change, adaptation occurs, and a regular slow rhythm of discharge is re-established. Such adapted frequencies plotted against temperature yield the familiar type of temperature curve with a positive temperature coefficient.

Lateral-line and stretch receptors of *Raja* show a similar paradoxical behaviour with respect to temperature changes, but on an entirely different scale. Stimuli of the order of 10° are required to produce an effect, and their maximal responses, compared with those of the ampullae, are minute. This is interpreted as meaning that the paradoxical effect of temperature changes is common to sensory rhythms in general, and possibly to some tonic muscles as well, but that in the ampullae of Lorenzini the mechanism has become specialized and sensitized for the performance of a specific thermosensory function.

The general form of a hypothesis to account for the paradoxical temperature effect is suggested. It involves the assumption of underlying reactions with different time factors of kinetic equilibration.

No cutaneous thermosensory endings have been found in rays, and it is concluded that the ampullae of Lorenzini are the only temperature receptors of elasmobranch fishes.

An interferometric wave-length comparison of the red cadmium radiation emitted by different sources. By D. V. Gogate and W. Ewart Williams, *King's College, London.* (*Communicated by O. W. Richardson, F.R.S.—Received* 5 *May* 1938.)

A comparison is made between the wave-lengths of the red cadmium line ($\lambda = 6438$ A) emitted by a G.E.C. Osira cadmium lamp and a Schüler hollow cathode discharge by photographing the Fabry-Pérot fringes due to the two sources side by side in the same horizontal plane by means of a specially designed plate holder. Interferometer gaps up to 10 cm. were used and the greatest difference recorded was less than 1 part in 100×10^6 or 0·00006 A.

ABSTRACTS

OF PAPERS COMMUNICATED TO THE ROYAL SOCIETY OF LONDON

In accordance with a resolution of Council, summaries or abstracts of papers are to be published as soon as practicable. The publication of such abstracts in no way indicates that the papers have been accepted for publication in any fuller form. These abstracts will be issued for convenience with the "Proceedings of the Royal Society of London" but do not form a part of the "Proceedings".

---

16 JUNE 1938

Sorption of polar and non-polar gases by zeolites. Parts I and II.

By RICHARD M. BARRER. (*Communicated by E. K. Rideal, F.R.S.—Received 2 May* 1938.)

The sorption on five different zeolites representative of the three types of zeolite structure—three-dimensional networks, fibrous, and laminar—has been studied for helium, hydrogen, argon, nitrogen, and ammonia. The more simple types of equilibria obtained with the permanent gases were used to interpret the more complex equilibria observed with ammonia. In the case of the permanent gases, only the three-dimensional networks (chabasite and analcite) sorbed gas within the dehydrated lattice. The fibrous and platy zeolites underwent a reorganization of ions around the water-bearing channels which was sufficient to exclude even helium (diameter 2·5 A). How marked the change was, was indicated by calculations of the potential within one of these zeolites (natrolite) which showed that even if methane (4·16 A diameter) were to replace water in an *otherwise unchanged* natrolite it would do so with a high sorption potential. The sorption of ammonia, however, occurred in fibrous and platy zeolites as well as in those of three-dimensional network type. A complete gradation in behaviour was found in the series:

NH_3—chabasite, NH_3—analcite, NH_3—heulandite, NH_3—natrolite,

the first giving a simple sorption system, and the last an ammoniate, the other two being intermediate. It was observed that a new phase tended to appear when the act of sorption required a reorganization of the lattice in order to accommodate the molecule sorbed. It was possible, however, so to alter the zeolite by outgassing, that not even ammonia would penetrate (NH_3—scolecite); and also to retard the resorption rate greatly by a prolonged heating while not completely inhibiting it (NH_3—heulandite). Kinetic studies of sorption showed that two types of ammonia uptake occurred: (1) A diffusion of NH_3 down a concentration gradient (NH_3 in three-dimensional networks). (2) A diffusion of ammoniate-anhydrous zeolite interfaces (NH_3—natrolite); this latter process was auto-catalytic in its initial stages. Sorption velocities were most rapid in chabasite and analcite—in the latter instance contrary to Tiselius' earlier finding.

The sorption heats were measured for

NH_3, H_2, Ar, N_2 —chabasite,
NH_3, He, H_2, Ar, N_2—analcite,
NH_3, Ar —natrolite,
NH_3, H_2, Ar —scolecite,
NH_3 —heulandite,

in many instances as a function of both quantity of gas sorbed and of temperature. The temperature coefficient of the heats for the permanent gases was negligible, and the heats tended to diminish with rising temperature for ammonia, and to diminish with increasing amount sorbed for all the gases.

The apparent saturation maxima diminished nearly linearly with the absolute temperature for all the gases.

The results were compared and contrasted with some previous data.

On the stability of a viscous liquid between rotating coaxial cylinders.

By J. L. SYNGE. (*Communicated by S. Goldstein, F.R.S.—Received 5 May 1938.*)

For a small disturbance (independent of the azimuthal angle Φ) of the steady motion of a liquid between rotating coaxial cylinders, the equations of motion for cylindrical co-ordinates r, Φ, z may be written

$$\left(\frac{\partial}{\partial\tau}-\Phi\right)\Phi\psi=-2R\left(a+\frac{b}{\xi^2}\right)r_1\frac{\partial\eta}{\partial\zeta},$$

$$r_1\left(\frac{\partial}{\partial\tau}-\Phi\right)\eta=2Ra\frac{\partial\psi}{\partial\zeta},$$

$$\Phi=\frac{\partial^2}{\partial\xi^2}+\frac{1}{\xi}\frac{\partial}{\partial\xi}+\frac{\partial^2}{\partial\zeta^2}-\frac{1}{\xi^2},$$

where r_1, r_2 are the radii of the cylinders ($r_2>r_1$), ω_1, ω_2 their angular velocities ($\omega_1>0$), ν the kinematical viscosity,

$$a=\frac{1}{\omega_1}\frac{\omega_2r_2^2-\omega_1r_1^2}{r_2^2-r_1^2},\quad b=-\frac{1}{\omega_1}\frac{r_2^2(\omega_2-\omega_1)}{r_2^2-r_1^2},\quad R=\frac{\omega_1r_1^2}{\nu},$$

$$\tau=\nu t/r_1^2,\quad \xi=r/r_1,\quad \zeta=z/r_1.$$

η is the velocity of disturbance in the Φ-direction and ψ a function in terms of which the other two components of velocity are determined by means of the equation of continuity. On substituting

$$\psi=r_1f(\xi)e^{\sigma\tau+i\lambda\zeta},\quad \eta=ig(\xi)e^{\sigma\tau+i\lambda\zeta},$$

we obtain the ordinary differential equations

$$(L-\lambda^2-\sigma)(L-\lambda^2)f=-2\lambda R(a+b/\xi^2)g,$$

$$(L-\lambda^2-\sigma)g=-2\lambda Raf,$$

$$L=\frac{d^2}{d\xi^2}+\frac{1}{\xi}\frac{d}{d\xi}-\frac{1}{\xi^2},$$

with the boundary conditions $f=df/d\xi=g=0$ for $\xi=1$ and $\xi=\alpha$.

The first of these equations is multiplied by $\xi\bar{f}d\xi$ (the bar denoting the complex conjugate) and integrated over the range: the second equation is multiplied by $-\left(1+\frac{b}{a\xi^2}\right)\xi\bar{g}d\xi$ and integrated over the range. When the two resulting equations and their complex conjugates are added together, the imaginary part of σ disappears, and an expression is obtained for the real part of σ in terms of certain integrals. If $\omega_2 r_2^2>\omega_1 r_1^2$, it is seen that the real part of σ is negative, and hence there is stability under this condition. This is in agreement with the theory and experiments of G. I. Taylor, but his detailed theory dealt only with the case where the distance between the cylinders is small. The present result is quite general, and is obtained without heavy calculations.

The approximate prediction of skin friction and lift.

By N. A. V. Piercy, J. H. Preston and L. G. Whitehead. (*Communicated by E. F. Relf, F.R.S.—Received* 6 *May* 1938.)

Convenient methods are developed for predicting by successive approximation the laminar skin friction of cylinders according to any analytical assumption regarding the irrotational flow outside their boundary layers. These methods almost eliminate the need for computation till the stage is reached when remaining mathematical errors become small. The particularly simple third approximation to the solution of the boundary layer equations suffices for most purposes until breakaway is closely approached. Explicit formulae are obtained for the position of breakaway and for the skin friction of elliptic cylinders.

An appropriate assumption regarding the external flow, avoiding appeal to experiment, is derived by recognizing the existence of a wake to the extent of elongating the cylindrical profile from the points of breakaway to infinity. The theory of the modification is formulated and a rapidly convergent method of successive approximation developed for its application. It transpires that any reasonable assumption regarding the thickness and configuration of the rudimentary wake secures close agreement with experiment—in contrast to the familiar consequences of ignoring the wake.

A new criterion for determining the circulation round a lifting cylinder is proposed and, in illustration, the variation of lift with inclination of section is calculated from first principles for a given elliptic cylinder. The result agrees fairly with experiment, the "stall", in particular, being predicted with some accuracy.

Compared with earlier methods, those now advanced promise improved physical accuracy in the theoretical determination of skin friction, breakaway, lift and stall. Their analytical convenience will, it is hoped, facilitate investigations of great complexity.

Effusion phenomena in relativistic quantum statistics. By B. N. Srivastava. (*Communicated by M. N. Saha, F.R.S.—Received* 9 *May* 1938.)

In this paper the effusion phenomena in non-degenerate and degenerate matter have been investigated from relativistic mechanics. Matter obeying either Fermi-Dirac statistics or Bose-Einstein statistics has been considered. In § 1 non-degenerate matter has been considered, and the number of particles as well as the mass of matter effusing out has been calculated. The non-relativistic expressions are obtained as a limiting case. For $f = 10$ the departure from the value for a classical perfect gas amounts only to about 1 %. In § 2 the effusion has been calculated for a degenerate gas, degenerate in the sense of Fermi-Dirac statistics, both to a first approximation and to a second approximation. The non-relativistic formulae are obtained as a limiting case. In this case it is found that to a first approximation the effusion varies as $p^{\frac{4}{5}}$ and is independent of T, but on working to a second approximation a slight dependence on temperature is observed. The average mass of particles in the effusion stream is calculated.

The diffraction of X-rays by age-hardening aluminium copper alloys. By G. D. Preston. (*Communicated by P. M. S. Blackett, F.R.S.—Received* 10 *May* 1938.)

The paper contains a description of an effect observed in Laue and oscillating crystal photographs of an alloy of high purity aluminium with 4 % copper aged at room temperature and at 150° C. The effect is ascribed to the separation of copper on (100) planes of the aluminium lattice.

Rotatory dispersion in the amine series. Part IV. The optical activity of diamines. By W. C. G. Baldwin, M.A., Ph.D. (*Communicated by R. G. W. Norrish, F.R.S.—Received* 10 *May* 1938.)

Variations in Drude equations for the rotatory dispersions, and changes in specific rotation, which occur on neutralization, are utilized for the detection of induced dissymmetry in the amino-radical of optically active bases. When the lone pair of electrons of the nitrogen atom becomes bound in any way, the partial rotation associated with it disappears. An attempt is made to explain the behaviour of a few exceptional bases whose rotatory powers contain no contribution which can be attributed directly to the lone pair of electrons. Such bases do not give the large alteration in specific rotation which is usually produced by ionization.

The rotatory dispersion of crystalline ethylene-diamine sulphate is compared with that of l-propylene-diamine, l-diamino-*cyclo*hexane, and of their salts in solution. Measurements and analyses are given for these two optically active bases, as well as for their mono- and dihydrochlorides. The results are interpreted in terms of Boys's molecular model, and the absolute configuration of l-propylene-diamine is defined. In spite of statements in the literature to the contrary, the sign of the specific rotation of this base is actually reversed on ionization, and this experimental fact constitutes strong evidence in favour of the proposed model.

It is shown that the major contribution to the optical activity of l-propylene-diamine may be attributed to the absorption bands associated with the two lone pairs of electrons in the molecule. Assuming the presence of an unchanging dextrorotatory nucleus common to all three ionic forms, an attempt to estimate the other contributions to the rotatory power and dispersion is made.

The absorption bands between 2300 A and 2000 A which are without exception common to all amines previously investigated are in the cases of l-propylene-diamine and l-*cyclo*hexane-diamine not optically active. The induced dissymmetry is centred at wave-lengths between 1900 A and 1400 A, and is associated with two groups of absorption bands in this region, one due to the lone pairs of electrons, and the other to the more saturated CN and NH linkages.

Investigations of infra-red spectra absorption of the CH_2 group in the region of 3μ. By J. J. Fox and A. E. Martin. (*Communicated by Sir Robert Robertson, F.R.S.—Received* 11 *May* 1938.)

The absorption of seven substances containing CH_2 groups has been investigated in the region of 3μ. Besides the aromatic CH bands found about $3{\cdot}27\mu$ when a benzene ring is present, two or more bands near $3{\cdot}4\mu$ are attributable to the CH_2 group.

A critical survey of the data available for ethylene enables us to explain the bands observed for CH_2 groups in more complicated molecules. When only one CH_2 group is contained in the molecule two bands are observed and these are identified as the CH valency vibrations in and out of phase, but when the number of CH_2 groups is increased, the coupling between them may increase the number of bands as in diozan, for example, where four strong bands are observed.

A satisfactory approximation to the molecules studied by us is found in a model of the formaldehyde type in which the CH_2 group is attached to a mass of similar magnitude. This model enables us to explain the observed frequencies provided that the CH force constant is $4{\cdot}52 \times 10^5$ dynes/cm., a value considerably lower than $5{\cdot}02 \times 10^5$ for ethylene and $5{\cdot}88 \times 10^5$ for acetylene. The important result is obtained that the mass to which the CH_2 group is attached and the various other force constants involved in compounds of the type investigated have very little effect on the positions of the CH_2 absorption bands in the 3μ region, except in rare cases involving strain of the bonds connecting the CH_2 groups to the rest of the molecule, e.g. fluorine.

The refraction and dispersion of air for the visible spectrum. By J. E. Sears, Junr. and H. Barrell. (*Communicated by W. L. Bragg, F.R.S.—Received* 11 *May* 1938.)

The refraction and dispersion of dry, CO_2-free air in the visible spectrum has been investigated, at temperatures between 12 and 31° C. and at pressures between 100 and 800 mm. of mercury, by means of an interference refractometer which employs Brewster's fringes produced by two 67 cm. Fabry-Pérot étalons. The results are used to provide a general interpolation equation from which definitive values of the refractive index may be calculated; the magnitude of the probable difference between a single observed value of refractive index and the value calculated from the equation is $\pm 0{\cdot}01_5 \times 10^{-6}$. The form of this equation, being derived from considerations based on the Lorenz-Lorentz relation between refractive index and density of transparent media and on D. Berthelot's representation of the characteristic behaviour of gases at low pressures, enables interesting comparisons to be made between the results of refractivity and compressibility measurements on air regarding the departure of the gas from Avogadro's law.

The effect of humidity on the refraction of atmospheric air has been determined from refractivity measurements on moist, CO_2-free air containing known amounts of unsaturated vapour.

No evidence has been found in favour of Tilton's suggested relationship between the refractivity of air and sunspot activity.

On the blood circulation and metabolism of Sabellids. By H. MUNRO FOX, F.R.S. (*Received* 13 *May* 1938.)

Most of the blood vessels of sabellid worms are rhythmically contractile. The rates of contraction increase in the following order: crown capillaries, lateral vessels, ventral vessel, peri-intestinal sinus. The rate of contraction is more rapid in small than in large individuals.

There is evidence that the movement of the blood is a circulation.

When *Sabella* retires into its tube the contractions of the peri-intestinal sinus stop just before those of the crown vessels, while the lateral vessels do not cease contracting until later. The oxychlorocruorin in the stagnant blood is not completely reduced.

There are capillary blood vessels supplying body-wall glands but none in the muscles. The muscles must therefore obtain oxygen from the sea water and from the coelomic fluid, which is richly supplied with blood capillaries. The high total oxygen capacity of the blood and the high unloading tension of chlorocruorin may be correlated with this indirect oxygen supply to the muscles from the blood through the coelomic fluid.

Spirographis is negatively thigmotropic, negatively geotropic and positively phototropic. These tropisms involve permanent curvatures of the tube which are probably accomplished by softening the tube through digestion, together with secretion of elastic tube substance.

Spirographis deprived of its tube will live indefinitely in aerated water, but dies in water deficient in oxygen, although worms with their tubes remain healthy in this water.

Spirographis deprived of its tube and put into a glass tube, lines the glass tube with a secretion and lengthens it like a normal tube. Worms in these glass tubes remain healthy in water deficient in oxygen.

It can be seen from *Spirographis* in glass tubes that the worms in their tubes execute vigorous rhythmic body movements. These movements are unaffected by oxygen deficiency or carbon dioxide excess in the water.

The rhythmic body movements probably continually renew the water in the tube. These movements are feeble or absent in tubeless worms. It is probably for this reason that tubeless worms die in water deficient in oxygen, while worms with their tubes can live in water deficient in oxygen by continually renewing the water in contact with the body. Tubeless worms are viable in aerated water probably because they can obtain enough oxygen without body movements. These movements must be necessary to worms with their tubes, even when the surrounding water is aerated, because of the small lumen of the tube and the difficulty of renewing water through the anterior opening of the tube alone.

The tube has thus a second function in addition to being protective; it provides the mechanical stimulus for reflex rhythmic body movements necessary to the animal with its tube.

When *Spirographis* is removed from its tube, its oxygen consumption falls to three-quarters of the initial value on the same day, and to one-half on the following day, while its ammonia excretion falls on the day of removal from the tube to four-fifths of the initial value. This fall in metabolism must be due in part to the absence of vigorous rhythmic body movements in the tubeless worms.

When the crown of *Spirographis* is amputated, the rate of oxygen consumption of the worm falls to one-third of the value for the intact worm. This is true both of worms with their tubes and of those deprived of tubes.

The rate of oxygen consumption of the isolated crown of *Sabella* is 2½ times that of the whole worm, but this high rate of metabolism of the crown is not sufficient to account for the fall in oxygen consumption of the worm after the crown has been amputated. The crown is therefore a respiratory organ.

During the regeneration of fragments of *Sabella* the oxygen consumption slowly increases from a low initial value to that of the intact tubeless worm.

An X-ray investigation of slowly cooled copper-nickel-aluminium alloys. By A. J. Bradley and H. Lipson. (*Communicated by W. L. Bragg, F.R.S.—Received* 13 *May* 1938.)

From an X-ray examination of slowly cooled CuNiAl alloys, a phase diagram has been constructed. The boundaries of the single-phase, two-phase and three-phase fields have been fixed from powder photographs. A novel feature of the work is the use made of the relative intensities of the X-ray reflexions to estimate the proportions of the constituents in two- and three-phase alloys.

The single-phase fields are derived from the copper-aluminium and nickel-aluminium systems, with the exception of (Cu_3NiAl_6), which has a new type of structure, derived from a deformed body-centred cube.

The molecular structures of carbon and silicon tetrafluorides. By C. R. Bailey, J. B. Hale and J. W. Thompson. (*Communicated by C. K. Ingold, F.R.S.—Received* 13 *May* 1938.)

The infra-red absorption spectra of the vapours of CF_4 and SiF_4 have been examined with a prism spectrometer between 1 and 18μ. The results are compared with those for the Raman effect as obtained by other workers. Certain molecular constants are determined, and it is found that a force field of the type devised by Urey and Bradley fits the observations very closely indeed. The internal electrical polarity of such tetrahedral pentatomic molecules is discussed with particular reference to the type of bonding, and to the interatomic distances as obtained by different methods.

The measurement of sand storms.

By R. A. Bagnold. (*Communicated by G. I. Taylor, F.R.S.—Received* 16 *May* 1938.)

Measurements are described of wind velocity and sand movement during storms in the Egyptian Desert. The results are shown to confirm in the following respects the author's previous laboratory (wind tunnel) work:

(*a*) *The modification of the normal wind velocity distribution caused by the dr*[illegible] *of the surface sand movement.*

(*b*) The relation between the rate of flow of sand over a dune and the wind gra[illegible] required to drive it.

(*c*) The particular form taken by the grading diagram of sand which has been picked up by the wind and is in process of transport over the surface.

A special multi-tube manometer is described which was designed for operation on the open surface of a dune during a sand storm.

Selective neutralization of the luteinizing activity of gonadotrophic extracts by antisera.

By I. W. Rowlands. (*Communicated by A. S. Parkes, F.R.S.—Received* 16 *May* 1938.)

Experiments on immature female rats show that it is possible to change the qualitative activity of gonadotrophic extracts of the pituitary by the simultaneous administration of anti-gonadotrophic sera. Two extracts of horse pituitary were used, both of which alone produced large numbers of corpora lutea in the immature ovary. Two antisera were prepared by injecting batches of rabbits daily for periods of 3 and 5 months respectively with gonadotrophic extracts of ox pituitary.

It was found that with suitable doses of antiserum the luteinizing activity of the extract was completely inhibited, although the response measured by increase in weight of the ovary was only slightly altered. Histologically, the response was one of pure follicle stimulation, as shown by the presence of numerous large follicles, and the absence of corpora lutea or any luteinization of the membrana granulosa. The uterus of all these animals was greatly distended with fluid such as is found in the normal animal during oestrus. Since these effects were not produced by variation in the dosage of horse pituitary extract alone, they cannot be due to a partial quantitative neutralization of the extract by the serum. When larger amounts of antiserum than the optimum were given there was a progressively increasing inhibition of follicular stimulation and of the ovarian weight response.

The significance of these results is discussed in relation to the probability that separate follicle stimulating and luteinizing hormones are present in pituitary extracts.